BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14565 A1

(51) International Patent Classification[7]: C12N 15/53, 15/63

(21) International Application Number: PCT/AU00/00966

(22) International Filing Date: 11 August 2000 (11.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PQ 2156 11 August 1999 (11.08.1999) AU

(71) Applicants *(for all designated States except US)*: **THE UNIVERSITY OF QUEENSLAND** [AU/AU]; St Lucia, QLD 4072 (AU). **VANDERBILT UNIVERSITY** [US/US]; Nashville, TN 37232 (US). **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US]; Pasadena, CA 91125 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GILLAM, Elizabeth, J.** [AU/AU]; 126 Market Street, Indooroopilly, QLD 4069 (AU). **NOTLEY, Lisa, M.** [AU/AU]; Unit 5, 34 Cordeaux Street, West End, QLD 4101 (AU). **DEVOSS, James** [AU/AU]; 23 Prospect Terrace, St Lucia, QLD 4067 (AU). **GUENGERICH, F., Peter** [US/US]; Department Of Biochemistry And Centre In Molecular Toxicology, Vanderbilt University School of Medicine, 638 Medical Research Building I, 23rd Avenue at Pierce, Nashville, TN 37232-0146 (US). **VOLKOV, Alexander** [US/US]; Genencor International Inc, 925 Page Mill Road, Palo Alto, CA 94304/1013 (US).

(74) Agents: **HUGHES, E., John, L.** et al.; Davies Collison Cave, Level 3, 303 Coronation Drive, Milton, QLD 4064 (AU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/14565 A1

(54) Title: PIGMENT PRODUCTION BY CELLS HAVING INTRODUCED CYTOCHROME P450 SEQUENCES

(57) Abstract: The present invention relates generally to a method for producing a pigment in or by biological material including plant, animal and microbial cells. More particularly, the present invention is directed to the use of a cytochrome P450 protein or a derivative or homologue thereof together with one or more homologous or heterologous associated proteins to facilitate the conversion of a substrate, to a coloured product. The method of the present invention is predicated in part on the oxidation of indole or related molecules by the cytochrome P450 protein and a homologous or heterologous associated protein to indoxyl and oxidized indoxyl or related molecules which alone or in combination with other molecules can form a range of pigments including indigo and indirubin.

Pigment production by cells having introduced cytochrome P450 sequences.

FIELD OF THE INVENTION

The present invention relates generally to a method for producing a pigment in or by biological material including plant, animal and microbial cells. Particularly useful biological material contemplated herein includes microbial cells, plant cells such as plant tissue and organs including petals, flowers, stems, leaves and seeds, plant fibrous material such as cotton, and non-cellular plant material. More particularly, the present invention is directed to the use of a cytochrome P450 protein or a derivative or homologue thereof together with one or more homologous or heterologous associated proteins to facilitate the conversion of a substrate, when available to or in a cell, to a coloured product. The coloured product makes the cell in which it is produced, as well as surrounding tissue or medium to which the product leaks and tissue and plant material connected by a vascular system to the cell in which the product is formed, change colour. The present invention is further directed to microorganisms, transgenic animal cells and transgenic plants, parts of plants including cells of plants and progeny of said plants which are capable of expressing genetic material encoding a cytochrome P450 protein or a derivative or homologue thereof and one or more homologous or heterologous associated proteins wherein microorganisms, transgenic animal cells and plant cells expressing such genetic material or surrounding cells or cells and plant material connected at the vascular level to said first mentioned cells undergo a colour change relative to cells or plant tissue which do not express this genetic material. The present invention further provides parts of plants including cells, tissues and organs such as petals, flowers (e.g. cut or severed flowers), stems, leaves and seeds, fibrous material such as cotton and non-cellular material which have undergone a colour change resulting from a pigment whose production is facilitated by the presence of a cytochrome P450 protein or a derivative or homologue thereof and one or more homologous or heterologous associated proteins. The method of the present invention is predicated in part on the oxidation of indole or related molecules by the cytochrome P450 protein and a homologous or heterologous associated protein to indoxyl and oxidized indoxyl or related molecules which alone or in combination with other molecules can form

a range of pigments including indigo and indirubin. The production of pigment in microbial cells, animal cells or plant cells is useful *inter alia* in the ornamental flower industry, as commercial tags and as sources of pigments for the dye industry. The pigment produced is also useful as a marker for transformation of microbial, plant and animal cells.

BACKGROUND OF THE INVENTION

Reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that this prior art forms part of the common general knowledge in Australia or any other country.

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description. Nucleotide and amino acid sequences are referred to by a sequence identifier, i.e. <400>1, <400>2, etc. A sequence listing is provided after the claims.

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the chemical, agricultural and horticultural industries. Of particular importance is the use of genetic techniques to introduce useful phenotypes in plants such as disease resistance, frost resistance and altered flower colour.

The ability to alter flower colour has been a major focus of the horticultural industry. The use of genetic manipulation has the potential to vastly increase the range of flower colour which has so far been achieved using conventional breeding practices. One approach has been to manipulate the flow of metabolites through the flavonoid biosynthetic pathway. This pathway results in the production of anthocyanins which are glycosylated derivatives of cyanidin, delphinidin, petunidin, peonidin, malvidin and pelargonidin. The flavonoid pathway is well established and the pattern of hydroxylation of the B-ring of anthocyanins plays a key role in determining petal colour (see International Patent Application No. PCT/AU94/00265 [WO 94/28140]).

In order to fully exploit the potential spectrum of colour change, other approaches of manipulating plant colour need to be considered.

Cytochrome P450 proteins are a superfamily of haemoproteins which catalyze the oxidation of a diverse range of chemicals in a wide variety of organisms. Cytochrome P450 proteins are involved, for example, in clearance of xenobiotic chemicals, biosynthesis of hormones and other signalling molecules and participate in homeostatic mechanisms. Cytochrome P450-dependent monooxygenases require for activity, associated proteins which catalyze the transfer of electrons from, for example, a co-enzyme to a prosthetic haem group on the cytochrome P450 molecule. Generally, the associated protein is a reductase such as a co-enzyme-dependent cytochrome P450 reductase. An associated protein may also have a role in facilitating interaction between a cytochrome P450 and a reductase. An example of such a molecule is cytochrome b_5 (De Vetten *et al.,* 1999). This may be used with a cytochrome P450 protein or in addition to another associated protein.

In work leading up to the present invention, the inventors observed that microorganisms engineered to express genetic material encoding a cytochrome P450 protein turned themselves and their culture medium blue. In accordance with the present invention, it is proposed to exploit this phenomenon to generate pigment production in microbial cells, animal cells and in plant cells and tissue and other plant material.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

One aspect of the present invention is directed to the use of genetic material encoding a cytochrome P450 protein or a functional derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material.

Another aspect of the present invention provides for a use of genetic material encoding a cytochrome P450 protein or a derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material wherein the genetic material, encoding the cytochrome P450 protein or its functional derivative or homologue, is co-expressed with genetic material encoding an associated protein.

A further aspect of the present invention is directed to the use of genetic material encoding a mammalian or microbial cytochrome P450 monooxygenase or a functional derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof and upon expression of said genetic material wherein said genetic material encoding the cytochrome P450 monooxygenase or its functional derivative or homologue is co-expressed with genetic material encoding an associated protein selected from a co-enzyme-dependent cytochrome P450 reductase and/or a cytochrome b_5 or a functional derivative or homologue of either molecule.

Still another aspect of the present invention provides a method of inducing a colour change in a plant or part of a plant said method comprising introducing into plant cells a genetic construct encoding a cytochrome P450 protein or a functional derivative or homologue thereof, regenerating a plant from said plant cells and subjecting said plant to appropriate conditions sufficient to permit expression of a nucleotide sequence on said genetic construct which encodes said cytochrome P450 protein such that in the presence of indole or a precursor, analogue or derivative thereof said cytochrome P450 protein facilitates the conversion of indole or its precursor, analogue or derivative to an intermediate which is capable of oxidation to a pigment.

Yet another aspect of the present invention provides a genetically modified cell or multicellular animal or plant or progeny thereof or parts of said transgenic plant or animal wherein said transgenic plant or animal comprises expressible genetic material encoding a microbial or mammalian cytochrome P450 protein.

Even yet another aspect of the present invention provides a genetically modified cell or multicellular animal or plant or progeny or parts thereof comprising expressible genetic material encoding a microbial or mammalian cytochrome P450 monooxygenase such that upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue or derivative is converted to an intermediate which is capable of oxidation to a pigment.

A further aspect of the present invention is directed to a transgenic plant or progeny or parts thereof comprising expressible genetic material encoding a microbial or mammalian cytochrome P450 monooxygenase such that upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue or derivative is converted to an intermediate which is capable of oxidation to a pigment.

Still another aspect of the present invention provides a transfected or transformed cell, tissue, organ or non-cellular material which contains or is capable of producing

cytochrome P450 protein or a functional derivative or homologue thereof.

Yet another of the present invention is further directed to flowers (e.g. cut or severed flowers), parts of plants, fibrous material from plants (e.g. cotton), non-cellular material and reproductive portions of plants (e.g. cuttings, seeds, pollen, callus) which are coloured or have the capacity to be coloured in the presence of indole or a precursor thereof due to the presence of cytochrome P450 protein (e.g. monooxygenase) and an associated protein (eg. NADPH-dependent cytochrome P450 reductase or cytochrome b_5).

Even yet another aspect of the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX_1(G/A) wherein X is any amino acid and X_1 is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

Still a further the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX_1(G/A) where X is any amino acid and X_1 is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

Yet another the present invention is directed to the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10 or <400>12;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4 or <400>6;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3 or <400>5; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Even yet another the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a

pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an NPR wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(ii) a nucleotide sequence defined by <400>13 and <400>15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13 and <400>15; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

A further aspect of the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(ii) a nucleotide sequence defined by <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Still another aspect of another aspect of the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an bacterial flavodoxin wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Yet another aspect of the present invention further contemplates the use of plant or plant material as a source of pigment for use, for example, in the dye or cosmetics industry or as commercial tags.

Even yet another aspect of the present invention further provides a genetic construct comprising a cistron encoding a microbial or mammalian cytochrome P450 protein for use, and more particularly, when used, in the generation of a transformed plant cell such that upon expression of said cistron, the cytochrome P450 facilitates conversion of indole or a precursor thereof to a molecule oxidizable to a pigment.

In a further aspect of the invention, the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

In still another aspect of the invention, the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

In yet another aspect of the present invention, the nucleotide sequence encoding the cytochrome P450 monooxygenase is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10, <400>12 or <400>21;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11 or <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4, <400>6 or <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3, <400>5 or <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

In even yet another aspect of the present invention, a particularly useful associated protein is human NPR (hNPR) encoded by:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>16;

(ii) a nucleotide sequence defined by <400> 15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

In a further aspect of the present invention, another useful associated protein is a bacterial flavodoxin encoded by:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagrammatic representation of the strategy used for generation of monocistronic and bicistronic constructs for co-expression of cytochrome P450 forms and hNPR. Solid bars represent P450 coding sequences; shading represents hNPR coding sequence; hatched bars represent other inserts, for example, linkers.

Figure 2 is a photographic representation of blue pigments in P450 2A6 cultures. (A) Cultures (2 ml) were prepared with (+) and without (-) isopropyl-β-D-thiogalactoside [IPTG] induction in TB media (48 hr at 29°C) with vigorous shaking. (B) Expression was induced by IPTG with (+) or without (-) the addition of 1.0 mM indole to the culture.

Figure 3 is a photographic representation of thin layer chromatography (TLC) separation of pigments formed by *E. coli* cells expressing DNA encoding cytochrome P450 2A6. The pigments (1,2,3,4) were separated by thin layer chromatography TLC as described in Example 3. The band ("O") at the TLC origin was subsequently shown by visible, ^{1}HNMR and mass spectrometry to be indigo, identical to compound 4 but not eluted due to the insoluble nature of the material.

Figure 4 is a photographic representation of spectra of pigments formed by *E. coli* cells expressing DNA encoding cytochrome 450 monooxygenase 2A6. Pigments (1,2,3,4) were separated by TLC as described and spectra of pigments 2 and 4 were recorded in $CHCl_3$. The spectrum of pigment 2 is characteristic of indirubin whereas pigment 4 corresponds to indigo.

Figure 5 is a graphical representation of ^{1}HNMR spectra of pigment 2 (indirubin)produced by cytochrome P450 monooxgenase 2A6 (Figure 3), at 400 MHz in d6-Me_2SO. Assignments are indicated. The connectivities and 2-D spectra are shown on the structure.

Figure 6 is a graphical representation of ^{1}HNMR spectra of pigment 4 (indigo) produced by cytochrome P450 monooxgenase 2A6 (Figure 3), at 400 MHz in d6-Me_2SO. Assignments are indicated. The connectivities and 2-D spectra are shown on the structure.

Figure 7 is a photographic representation of TLC showing *in vitro* production of blue pigments by bacterial membranes containing human cytochrome P450 protein and human NADPH-dependent cytochrome P450 reductase (hNPR). Each cytochrome P450 form is indicated. The arrow indicates the position of indigo. In the case of incubations containing P450 2A6, the indigo produced remained largely at the origin due to insolubility, a phenomenon also seen in Figure 3.

Figure 8 is a graphical representation showing relative *in vitro* production of blue pigments by bacterial membranes containing human cytochrome P450 protein and human NADPH-dependent cytochrome P450 reductase (hNPR). Each cytochrome P450 form is indicated. For cytochrome P450 monooxygenase 2C9, the various allelic variants are indicated as "*1", "*2", "*3". "HL16" is a human liver microsomal sample. In all cases, 0.8 mM indole was incubated with the membranes (0.2 μM P450) for 30 min at 37°C and dye production was analyzed by measuring absorbance at 665 nm.

Figure 9 is a representation of a timecourse showing the production of various metabolites from indole over time in incubations with recombinant P450 2A6 and hNPR in bacterial membranes. Experiments were performed as described in Example 6. A_{615} indicates total absorbance at 615 nm, indicating pigment production. tR 3.3 indicates a metabolite with retention time of 3.3 min under the HPLC conditions used. Other metabolites are indicated on the figure.

Figure 10 is a representation of the postulated reactions in cytochrome P450 monooxygenase-catalyzed pigment formation in *E. coli*.

Figure 11 is a diagrammatic representation showing postulated chemical mechanisms for cytochrome P450 monooxygenase-catalyzed pigment formation in *E. coli*.

Figure 12 is a diagrammatic representation showing the general strategy for subcloning cytochrome P450 and hNPR cDNAs into vectors for expression in plants. In the cases of P450 2A6 and hNPR, which contain an internal *Sac*I site, the final ligation of the ribosome binding site (RBS) coding sequence and *nos* terminator into p2.5GEM-II was done in two separate steps.

A summary of the sequence identifiers used in the subject specification is shown in Table 1.

TABLE 1

Summary of Sequence Identifiers

SEQUENCE IDENTIFIER	DESCRIPTION
<400>1	Nucleotide sequence encoding human derived cytochrome P450 2A6 monooxygenase for expression in bacterial cells
<400>2	Amino acid sequence of P450 2A6 encoded by <400>1
<400>3	Nucleotide sequence encoding human derived cytochrome P450 2E1 monooxygenase for expression in bacterial cells
<400>4	Amino acid sequence of P450 2E1 encoded by <400>3
<400>5	Nucleotide sequence encoding human derived cytochrome P450 2C19 monooxygenase for expression in bacterial cells
<400>6	Amino acid sequence of P450 2C19 encoded by <400>5
<400>7	Nucleotide sequence encoding human derived cytochrome P450 2A6 monooxygenase for expression in eukaryotic cells
<400>8	Amino acid sequence of P450 2A6 encoded by <400>7
<400>9	Nucleotide sequence encoding human derived cytochrome P450 2E1 monooxygenase for expression in eukaryotic cells
<400>10	Amino acid sequence of P450 2E1 encoded by <400>9
<400>11	Nucleotide sequence encoding human derived cytochrome P450 2C19 monooxygenase for expression in eukaryotic cells
<400>12	Amino acid sequence of P450 2C19 encoded by <400>11
<400>13	Nucleotide sequence encoding human derived cytochrome P450 reductase (NPR) for expression in bacterial cells
<400>14	Amino acid sequence of NPR encoded by <400>13

SEQUENCE IDENTIFIER	DESCRIPTION
<400>15	Nucleotide sequence of human derived NPR for expression in eukaryotic cells
<400>16	Amino acid sequence of NPR encoded by <400>15
<400>17	Nucleotide sequence of genetic construct for bacterial production of P450 2A6
<400>18	Nucleotide sequence of genetic construct for bacterial production fof P450 2E1
<400>19	Nucleotide sequence of genetic construct for bacterial product of P450 2C19
<400>20	Nucleotide sequence encoding P450 cin, a cytochrome P450 monooxygenase from *Citrobacter brakii*
<400>21	Amino acid sequence of P450 cin encoded by <400>20
<400>22	Nucleotide sequence encoding flavodoxin
<400>23	Amino acid sequence of flavodoxin encoded by <400>22

A summary of abreviations used in the subject specification is provided in Table 2.

TABLE 2

Abbreviations

ABBREVIATION	DESCRIPTION
P450 A26	Cytochrome P450 2A6 monooxygenase
P450 2E1	Cytochrome P450 2E1 monooxygenase
P450 2C19	Cytochrome P450 2C19 monooxygenase
P450 cin	Cytochrome P450 cin monooxygenase from *Citrobacter brakii*
IPTG	Isopropyl-β-D-thiogalactoside
TLC	Thin layer chromatography
NPR	NADPH-cytochrome P450 reductase
hNPR	Human-derived NADPH-cytochrome P450 reductase
Associated protein	A protein which catalyzes transfer of electrons from a co-enzyme to prosthetic haem group on cytochrome P450 protein or facilitates interaction of cytochrome P450 protein and a cytochrome P450 reductase

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Accordingly, one aspect of the present invention is directed to the use of genetic material encoding a cytochrome P450 protein or a functional derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material.

More particularly, the present invention provides for the use of genetic material encoding a cytochrome P450 protein or a functional derivative or homologue thereof in the generation of a transgenic plant which transgenic plant has the capacity to produce a pigment in all or part of its tissue or non-cellular material in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material.

The cytochrome P450 protein is preferably a cytochrome P450 monooxygenase and requires one or more associated proteins for activity. The associated protein may be a reductase such as a co-enzyme-dependent cytochrome P450 reductase and/or may be cytochrome b_5 or a homologue thereof. The associated protein may be a naturally occurring molecule in the cell into which the genetic material encoding the cytochrome P450 protein is introduced or it may be introduced simultaneously with, subsequently to or separately from the genetic material encoding the cytochrome P450 protein. In one embodiment, the genetic material comprises a bicistronic construct encoding both the cytochrome P450 protein and the associated protein. In another embodiment, the genetic material comprises two or more separate monocistronic constructs each encoding a cytochrome P450 protein and one or more associated proteins. In still yet another embodiment, the genetic material encodes a cytochrome P450 protein and its activity is facilitated by one or more host cell associated proteins. In even yet another aspect, the genetic material is a tricistronic or multicistronic construct encoding a cytochrome P450 protein and at least two associated proteins.

Another aspect of the present invention provides for a use of genetic material encoding a cytochrome P450 protein or a derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material wherein the genetic material, encoding the cytochrome P450 protein or its functional derivative or homologue, is co-expressed with genetic material encoding an associated protein.

More particularly, the present invention is directed to a use of genetic material encoding a cytochrome P450 protein or a derivative or homologue thereof in the generation of a transgenic plant which transgenic plant has the capacity to produce a pigment in all or part of its tissue or non-cellular material in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material wherein the genetic material encoding the cytochrome P450 protein or its functional derivative or homologue is co-expressed with genetic material encoding an associated protein. The cell may be a microbial cell (e.g. a bacterial, yeast or fungal cell), an animal cell (e.g. an insect, fish, amphibian, avian or mammalian cell) or a plant cell.

The associated protein may be indigenous to the cell or may have been previously introduced. The present invention covers both homologous and heterologous cytochrome P450 protein-associated protein combinations. A homologous combination is one which is naturally occurring. That is, the associated protein is one which is normally associated with the cytochrome P450 protein in a naturally occurring cell. A heterologous combination is one where the associated protein is not normally associated with the cytochrome P450 enzyme. An associated protein includes a molecule such as cytochrome b_5 or a functional homologue or derivative thereof. The term "co-expressed" includes co-expression on a single bicistronic, tricistronic or multicistronic construct, expression of separate monocistronic constructs and expression of one or more genetic construct together with expression of a nucleotide sequence contained in genomic DNA.

The preferred associated protein is a protein which can catalyze the transfer of electrons. One example of a reductase is a co-enzyme-dependent cytochrome P450 reductase. One particularly useful reductase in the practice of the present invention is an NADPH-cytochrome P450 reductase or "NPR" which catalyses the transfer of electrons from NADPH *via* FAD and/or FMN to the prosthetic haem group on the cytochrome P450 molecule. A particularly useful NPR is a human-derived NPR which is designated herein "Hnpr". Another useful reductase is a redoxin reductase. The preferred cytochrome P450 protein is a monooxygenase and is of mammalian or microbial origin. Particularly useful cytochrome P450 monooxygenases in the practice of the present invention are of human or primate origin. The present invention extends, however, to cytochrome P450 monooxygenases of bacterial or eukaryotic origin.

Another aspect of the present invention is directed to the use of genetic material encoding a mammalian or microbial cytochrome P450 monooxygenase or a functional derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof and upon expression of said genetic material wherein said genetic material encoding the cytochrome P450 monooxygenase or its functional derivative or homologue is co-expressed with genetic material encoding an associated protein selected from a co-enzyme-dependent cytochrome P450 reductase and/or a cytochrome b_5 or a functional derivative or homologue of either molecule.

More particularly, the present invention provides for the use of genetic material encoding a mammalian or microbial cytochrome P450 monooxygenase or a functional derivative or homologue thereof in the generation of a transgenic plant which transgenic plant has the capacity to produce a pigment in all or part of its tissue or non-cellular material in the presence of indole or a precursor, analogue or derivative thereof and upon expression of said genetic material wherein said genetic material encoding the cytochrome P450 monooxygenase or its functional derivative or homologue is co-expressed with genetic material encoding an associated protein selected from a co-enzyme-dependent cytochrome

P450 reductase and/or a cytochrome b_5 or a functional derivative or homologue of either molecule.

A preferred reductase is an NADPH-dependent cytochrome P450 reductase.

As stated above, the term "co-expression" includes expression of genetic material encoding both the cytochrome P450 monooxygenase and the associated protein in a bicistronic, tricistronic or multicistronic construct as well as expression of separate genetic constructs or co-expression of the cytochrome P450 monooxygenase genetic material with genetic material encoding an associated protein indigenous to the cell into which the genetic material encoding the cytochrome P450 monooxygenase is introduced.

The term "non-cellular material" in relation to plants includes vascular regions such as the phloem and xylem, fibrous material such as cotton and non-living regions of the plant.

Although not intending to limit the present invention to any one theory or mode of action, it is proposed that the cytochrome P450 protein and the associated protein(s) convert indole to indoxyl. This compound may then spontaneously oxidize in the presence of oxygen to oxidized indoxyl and then be spontaneously transformed to indigo. This process may also be catalyzed by the cytochrome P450 protein. Alternatively, indoxyl may be further transformed by cytochome P450 catalysis to indigo by a different mechanism (see Figure 11). Another product of cytochrome P450-dependent oxidation of indole is isatin. Another pigment, indirubin, is also generated when isatin combines with indoxyl. The pigments produced may remain in the cells, tissues or organs such as, in the case of plants, cells of petals, flowers, stems, leaves or seeds, may leak to surrounding cellular or non-cellular regions or may be transported to other regions *via* the plant vascular system. The pigment may also be present in fibrous material such as cotton or even non-cellular material. By expressing the genetic material encoding the cytochrome P450 protein and the associated protein with a developmentally or organ specific promoter, particular parts of the plant can be targeted for a colour change. The colour is generally dark, such as dark blue or black or may be a purple to red colour. Different shades of colours may be obtained

by providing chemically modified substrates for the cytochrome P450 monooxygenases or providing to the plant modified substrates or compounds which before or after cytochrome P450-dependent metabolism may complex with endogenous indole or a precursor, analogue, derivative or metabolites thereof. In addition, a range of colours may be obtained following interactions between indigo and other related dyes or their precursors and plant molecules associated with pigmentation such as flavonoid molecules.

All plants are encompassed by the present invention, i.e. monocotyledonous plants and dicotyledonous plants. Particularly useful plants are flower crop plants including rose, carnation, petunia, lisianthus, lily, iris, tulip, freesia, delphinium, limonium, pelargonium, commercial crops such as cotton and food crops (such as wheat, rice, barley and sugar cane).

Yet another aspect of the present invention provides a method of inducing a colour change in a plant or part of a plant, said method comprising introducing into plant cells a genetic construct encoding a cytochrome P450 protein or a functional derivative or homologue thereof, regenerating a plant from said plant cells and subjecting said plant to appropriate conditions sufficient to permit expression of a nucleotide sequence on said genetic construct which encodes said cytochrome P450 protein such that in the presence of indole or a precursor, analogue or derivative thereof said cytochrome P450 protein facilitates the conversion of indole or its precursor, analogue or derivative to an intermediate which is capable of oxidation to a pigment.

Preferably, the cytochrome P450 protein is a monooxygenase of microbial or mammalian origin. The term "microbial origin" includes bacterial origin. Reference to a "cytochrome P450 protein includes and encompasses hybrid molecules from two or more cytochrome P450 proteins as well as naturally occurring and artificially created mutants and derivatives.

The activity of the cytochrome P450 protein requires the presence of one or more accessory proteins which may be present in and be indigenous to the plant cells or may

also be introduced on a separate genetic construct or on the same genetic construct which encodes the cytochrome P450 protein.

In one aspect, the indole (or its precursor) covereted to indigo or its precursor is endogenous indole. However, increased levels of indole or its precursor, analogue or derivative thereof may be provided to the plant *via,* for example, its root system such as in a fertilizer. Alternatively, cut flowers or other severed portions of the plant may be placed in a solution of indole or precursor, analogue or derivative thereof. The vascular system, in the case of a plant, will then transport the indole into the plant tissue cells and pigment production will occur in the presence of the cytochrome P450 protein and the associated protein. In a further aspect of the invention, cells are engineered to express genetic material encoding a tryptophanase or another enzyme in order to promote and/or endogenous levels of indole or its precursors or homologues. Preferably, the genetic material encoding the tryptophanase or other enzymes would be able to be regulated at the developmental and/or organ and/or tissue specific levels.

This aspect of the present invention, insofar as it relates to plants, further extends to progeny of the plants engineered to express the cytochrome P450 protein as well as vegetative, propagative and reproductive parts of the plants (including cuttings, pollen, seeds and callus).

Another aspect of the present invention provides a genetically modified cell or multicellular animal or plant or progeny thereof or parts of said transgenic plant or animal wherein said transgenic plant or animal comprises expressible genetic material encoding a microbial or mammalian cytochrome P450 protein.

More particularly, the present invention provides a transgenic plant or progeny thereof or parts of said transgenic plant wherein said transgenic plant comprises expressible genetic material encoding a microbial or mammalian cytochrome P450 protein.

Preferably, the cytochrome P450 protein is a monooxygenase.

Even more particularly, the present invention provides a genetically modified cell or multicellular animal or plant or progeny or parts thereof comprising expressible genetic material encoding a microbial or mammalian cytochrome P450 monooxygenase such that upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue or derivative is converted to an intermediate which is capable of oxidation to a pigment.

Still more particularly, the present invention is directed to a transgenic plant or progeny or parts thereof comprising expressible genetic material encoding a microbial or mammalian cytochrome P450 monooxygenase such that upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue or derivative is converted to an intermediate which is capable of oxidation to a pigment.

The term "genetically modified" is used in its broadest sense and includes introducing gene(s) into cells, mutating gene(s) in cells and altering or modulating the regulation of gene(s) in cells.

A "part" of a plant includes flowers (e.g. cut or severed flowers), petals, stems, leaves and fibrous material such as cotton and vegetative, propagative and reproductive material such as cuttings, pollen, seeds and callus.

Generally, the coding region(s) on the genetic material are operably linked to a single or multiple promoters. In one embodiment, the genetic construct is a bicistronic construct under the control of (i.e. operably linked to) a single promoter where the cistrons encode both the cytochrome P450 protein and the associated protein. In another embodiment, the genetic construct is a tricistronic or multicistronic construct.

As used herein, the term "genetic material" refers to any single-stranded or double-stranded nucleic acid molecule which at least comprises deoxyribonucleotides and/or

ribonucleotides, including DNA (cDNA or genomic DNA), RNA, mRNA, or tRNA, amongst others. The combination of such molecules with non-nucleotide substituents derived from synthetic means or naturally-occurring sources is also contemplated by the present invention.

"Derivatives" of the genetic sequence of the invention refers to any isolated nucleic acid molecule which comprises at least 10 and preferably at least 20 contiguous nucleotides derived from the genetic sequence as described herein according to any embodiment. A derivative includes a part, fragment, portion or analogue. A derivative also includes a fusion molecule between two more genetic sequences encoding cytochrome P450 molecules.

An "analogue" of a genetic sequence of the invention means any isolated nucleic acid molecule which is substantially the same as a nucleic acid molecule of the present invention or its complementary nucleotide sequence as described herein according to any embodiment, notwithstanding the occurrence of any non-nucleotide constituents not normally present in said isolated nucleic acid molecule, for example carbohydrates, radiochemicals including radionucleotides, reporter molecules such as, but not limited to, alkaline phosphatase or horseradish peroxidase, amongst others. A "homologue" is a functionally similar molecule from a different species or strain.

Generally, analogues or derivatives of the nucleic acid molecule of the invention are produced by synthetic means or alternatively, derived from naturally-occurring sources. For example, the nucleotide sequence of the present invention may be subjected to mutagenesis to produce single or multiple nucleotide substitutions, deletions and/or insertions.

The genetic sequence of the present invention may comprise a sequence of nucleotides or be complementary to a sequence of nucleotides which comprise one or more of the following: a promoter sequence, a 5' non-coding region, a *cis*-regulatory region such as a functional binding site for transcriptional regulatory protein or translational regulatory

protein, an upstream activator sequence, an enhancer element, a silencer element, a TATA box motif, a CCAAT box motif, or an upstream open reading frame, transcriptional start site, translational start site, and/or nucleotide sequence which encodes a leader sequence. The genetic sequence also encodes the cytochrome P450 enzyme and optionally an associated protein (e.g. cytochrome P450 reductase and/or cytochrome b_5).

The term "5' non-coding region" is used herein in its broadest context to include all nucleotide sequences which are derived from the upstream region of an expressible gene, other than those sequences which encode amino acid residues which comprise the polypeptide product of said gene, wherein 5' non-coding region confers or activates or otherwise facilitates, at least in part, expression of the gene.

The term "gene" is used in its broadest context to include both a genomic DNA region corresponding to the gene as well as a cDNA sequence corresponding to exons or a recombinant molecule engineered to encode a functional form of a product.

As used herein, the term "*cis*-acting sequence" or "*cis*-regulatory region" or similar term shall be taken to mean any sequence of nucleotides which is derived from an expressible genetic sequence wherein the expression of the first genetic sequence is regulated, at least in part, by said sequence of nucleotides. Those skilled in the art will be aware that a *cis*-regulatory region may be capable of activating, silencing, enhancing, repressing or otherwise altering the level of expression and/or cell-type-specificity and/or developmental specificity of any structural gene sequence.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or environmental stimuli, or in a tissue-specific or cell-type-specific manner. A promoter is usually, but not necessarily, positioned upstream or 5', of a structural gene, the expression

of which it regulates. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2kb of the start site of transcription of the gene.

In the present context, the term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of a structural gene or other nucleic acid molecule, in a plant cell. Preferred promoters according to the invention may contain additional copies of one or more specific regulatory elements to further enhance expression in a cell, and/or to alter the timing of expression of a structural gene to which it is operably connected.

The term "operably connected" or "operably linked" in the present context means placing a structural gene (e.g. cytochrome P450 protein or associated protein) under the regulatory control of a promoter which then controls expression of the gene. Promoters and the like are generally positioned 5' (upstream) to the genes which they control. In the construction of heterologous promoter/structural gene combinations, it is generally preferred to position the genetic sequence or promoter at a distance from the gene transcription start site that is approximately the same as the distance between that genetic sequence or promoter and the gene it controls in its natural setting, i.e., the gene from which the genetic sequence or promoter is derived. As is known in the art, some variation in this distance can be accommodated without loss of function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting, i.e., the genes from which it is derived.

A further aspect of the present invention provides a transfected or transformed cell, tissue, organ or non-cellular material which contains or is capable of producing cytochrome P450 protein or a functional derivative or homologue thereof. Preferably, the cytochrome P450 protein is a cytochrome P450 monooxygenase. This material will also contain or be capable of producing an associated protein such as a co-enzyme-dependent cytochrome P450 reductase or cytochrome b_5 or a functional homologue or derivative of either.

The genetic construct(s) of the present invention may be introduced into a cell by various techniques known to those skilled in the art. The technique used may vary depending on the known successful techniques for that particular organism.

Techniques for introducing recombinant DNA into cells include, but are not limited to, transformation using $CaCl_2$ and variations thereof, direct DNA uptake into protoplasts, PEG-mediated uptake to protoplasts, microparticle bombardment, electroporation, microinjection of DNA, microparticle bombardment of tissue explants or cells, vacuum-infiltration of tissue with nucleic acid, and T-DNA-mediated transfer from *Agrobacterium* to the plant tissue.

For microparticle bombardment of cells, a microparticle is propelled into a cell to produce a transformed cell. Any suitable ballistic cell transformation methodology and apparatus can be used in performing the present invention. Exemplary apparatus and procedures are disclosed by Stomp *et al.* (U.S. Patent No. 5,122,466) and Sanford and Wolf (U.S. Patent No. 4,945,050). When using ballistic transformation procedures, the genetic construct may incorporate a plasmid capable of replicating in the cell to be transformed.

Examples of microparticles suitable for use in such systems include 0.1 to 10 μm and more particularly 10.5 to 5 μm tungsten or gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

Once introduced into cells such as plant tissue, the expression of a cytochrome P450 protein and/or associated protein may be assayed in a transient expression system or it may be determined after selection for stable integration within for example, the plant genome.

Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant generated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons,

hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g. apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g. cotyledon meristem and hypocotyl meristem).

The regenerated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed to give homozygous second generation (or T2) transformant, and the T2 plants further propagated through classical breeding techniques.

Intended recipient plants of the cytochrome P450 protein or its derivative or homologue and/or associated protein or proteins (e.g. a cytochrome P450 reductase or cytochrome b_5) may first be modified such that the level of available indole or its precursor, analogue or derivative thereof for conversion to pigments (e.g. indigo) is altered. Generally, but not necessarily, the level is increased. A range of mutations in amino acid biosynthesis or metabolic pathways may be induced to result in accumulation of indole or an indole precursor. Alternatively, the indole or its precursor, analogue or derivative thereof is provided exogenously such as *via* a fertilizer or *via* a solution applied to the plant or parts of the plant.

The present invention is further directed to flowers (e.g. cut or severed flowers), parts of plants, fibrous material from plants (e.g. cotton), non-cellular material and reproductive portions of plants (e.g. cuttings, seeds, pollen, callus) which are coloured or have the capacity to be coloured in the presence of indole or a precursor thereof due to the presence of cytochrome P450 protein (e.g. monooxygenase) and an associated protein (e.g. NADPH-dependent cytochrome P450 reductase or cytochrome b_5).

Reference herein to "mammalian" in relation to a cytochrome P450 protein or associated protein includes a human, primate, livestock animal (e.g. sheep, pig, cow, horse, donkey), laboratory test animal (e.g. rabbit, mouse, guinea pig, hamster), companion animal (e.g. cat, dog) or a captive wild animal. Preferred mammals are primates and humans. Particularly preferred mammals are humans. Reference to "bacterial" includes prokaryotes

such as *E. coli, Pseudomonas* sp, *Acinetobacter* sp, and *Citrobacter* sp; the term "microbial" includes bacteria but also includes eukaryotic cells such as yeast, fungi and animal cells, e.g. insect cells.

Any and all microbial and mammalian cytochrome P450 proteins and in particular cytochrome P450 monooxygenases are contemplated by the present invention provided that such monooxygenases can act on indole or its precursors, analogues or derivatives. Preferred cytochrome P450 proteins comprise the consensus amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position. More particularly, the cytochrome P450 monooxygenases comprise the consequence sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position. The cytochrome P450 monooxygenase may also comprise a hybrid from two or more cytochrome P450 proteins. The cytochrome P450 monooxygenases according to this aspect of the present invention are capable of converting indole or a precursor, analogue or derivative to an intermediate which is capable of oxidation to a pigment. Conveniently, a suitable cytochrome P450 monooxygenase having the above consensus sequence and, in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, provides a pigment or a precursor of a pigment, may be selected from Genbank. Alternatively, a Nelson website may be selected such as: http://drnelson.utmem.edu/CytochromeP450.html; http://drnelson.utmem.edu/p450apub192.html; and http://drnelson.utmem.edu/p450bpub237.html.

The consensus sequence is obtained by comparing the amino acid sequence of the haem binding region of P450 proteins.

Particularly preferred cytochrome P450 monooxygenases include the human-derived P450 2A6, P450 2E1 and P450 2C19 monooxygenases. The nucleotide sequences encoding

these monooxygenases are given in <400>1, <400>3 and <400>5, respectively (corresponding amino acid sequences are given in <400>2, <400>4 and <400>6, respectively). These sequences are specifically designed for expression in bacterial cells. The nucleotide sequence given in <400>7, <400>9 and <400>11 (with corresponding amino acid sequences given in <400>8, <400>10 and <400>12, respectively) are respectively for P450 2A6, P450 2E1 and P450 2C19 when in a form suitable for expression in eukaryotic cells such as plant cells.

Accordingly, the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

More particularly, the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

Even more particularly, the present invention is directed to the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10 or <400>12;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4 or <400>6;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3 or <400>5; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

The cytochrome P450 monooxygenase of the present invention includes modified forms of the enzyme such as those resulting from site-directed or random mutagenesis as well as naturally occurring mutant forms.

The cytochrome P450 monooxygenase requires an associated protein such as NADPH-dependent cytochrome P450 (NPR) or cytochrome b_5. One particularly useful associated

protein is NPR. The nucleotide and corresponding amino acid sequences for an NPR from human cells are given in <400>13 and <400>14 (for use in bacteria) and <400>15 and <400>16 (for use in plants).

Accordingly, the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an NPR wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(ii) a nucleotide sequence defined by <400>13 and <400>15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13 and <400>15; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Another useful cytochrome P450 monooxygenase is of bacterial origin such as from *Citrobacter* sp. One particularly useful cytochrome P450 monooxygenase is from *Citrobacter brakii.* The nucleotide sequence and corresponding amino acid sequence is shown in <400>20 and <400>21, respectively. Its corresponding associated protein is a flavodoxin. The nucleotide sequence encoding flavodoxin and its corresponding amino acid sequence is shown in <400>22 and <400>23, respectively.

Accordingly, the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(ii) a nucleotide sequence defined by <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

In accordance with this embodiment, the bacterial cytochrome P450 monooxygenase is preferably operable in the presence of a flavodoxin or other suitably associated protein. Accordingly, another aspect of the present invention provides for the use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an bacterial flavodoxin wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or

conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. In a particularly preferred embodiment, nucleotide and sequence comparisons are made at the level of identity rather than similarity.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence similarity", "sequence identity", "percentage of sequence similarity", "percentage of sequence identity", "substantially similar" and "substantial identity". A "reference sequence" is at least 12 but frequently 15 to 18 and often at least 25 or above, such as 30 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (i.e. only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window" refers to a conceptual segment of typically 12 contiguous residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (i.e. gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerised implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (i.e. resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.* (1997). A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.* (1998).

The terms "sequence similarity" and "sequence identity" as used herein refers to the extent that sequences are identical or functionally or structurally similar on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity", for example, is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g. A, T, C, G, I) or the identical amino acid residue (e.g. Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, "sequence identity" will be understood to mean the "match percentage" calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using standard defaults as used in the reference manual accompanying the software. Similar comments apply in relation to sequence similarity.

Reference herein to a low stringency includes and encompasses from at least about 0 to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridization, and at least about 1 M to at least about 2 M salt for washing conditions. Generally, low stringency is at from about 25-30°C to about 42°C. The temperature may be altered and higher temperatures used to replace formamide and/or to give alternative stringency conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridization, and at least about 0.5 M to at least about 0.9 M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01 M to at least about 0.15 M salt for hybridization, and at least about 0.01 M to at least about 0.15 M salt for washing conditions. In general, washing is carried out $T_m = 69.3 + 0.41$ (G+C)% (Marmur and Doty, 1962). However, the T_m of a duplex DNA decreases by 1°C with every increase

of 1% in the number of mismatch base pairs (Bonner and Laskey, 1974). Formamide is optional in these hybridization conditions. Accordingly, particularly preferred levels of stringency are defined as follows: low stringency is 6 x SSC buffer, 0.1% w/v SDS at 25-42°C; a moderate stringency is 2 x SSC buffer, 0.1% w/v SDS at a temperature in the range 20°C to 65°C; high stringency is 0.1 x SSC buffer, 0.1% w/v SDS at a temperature of at least 65°C.

The present invention further contemplates the use of plant or plant material as a source of pigment for use, for example, in the dye or cosmetics industry, as commercial tags and as markers for transformation of microbial, plant and/or animal cells. The present invention further provides a genetic construct comprising a cistron encoding a microbial or mammalian cytochrome P450 protein for use, and more particularly, when used, in the generation of a transformed plant cell such that upon expression of said cistron, the cytochrome P450 facilitates conversion of indole or a precursor thereof to a molecule oxidizable to a pigment. Preferably, the cytochrome P450 protein is a mammalian or bacterial cytochrome P450 monooxygenase.

Preferably, the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

More preferably, the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

Even more preferably, the nucleotide sequence encoding the cytochrome P450 monooxygenase is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10, <400>12 or <400>21;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11 or <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4, <400>6 or <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3, <400>5 or <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Preferably, the genetic construct comprises another cistron encoding an associated protein such as but not limited to a co-enzyme-dependent cytochrome P450 reductase (NPR) and/or cytochrome b_5 or a functional derivative or homologue thereof.

A particularly useful associated protein is human NPR (hNPR) encoded by:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>16;

(ii) a nucleotide sequence defined by <400> 15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

Another useful associated protein is a bacterial flavodoxin encoded by:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

The above nucleotide sequences may encode the amino acid sequence of the naturally occurring molecule or a modified form such as resulting from site directed mutagenesis, random mutagenesis as well as naturally occurring mutant forms.

Another aspect of the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence encoding or complementary to a sequence encoding a cytochrome P450 cin monooxygenase from *Citrobacter brakii*.

Preferably, the cytochrome P450 monooxygenase comprises an amino acid sequence substantially as set forth in <400>21 or an amino acid sequence having at least about 50% similarity thereto after optimal alignment.

Preferably, the cytochrome P450 monooxygenase is encoded by a nucleotide sequence substantially as set forht in <400>20 or a nucleotide sequence having at least about 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

Aspects of the subject invention were made with United States Government Support under grant number CA44353 awarded by the National Institutes of Health. The United States Government may have certain rights in aspects of the instant invention. A licence for foreign filing for various aspects of the present invention has been obtained under Serial Number P-101,005.

The present invention is further described by the following non-limiting Examples.

EXAMPLE 1

Bacterial cultures

Bicistronic plasmids for expression of P450s and NADPH-P450 reductase were constructed as previously described (Parikh *et al.*, 1997). P450 2A6, P450 2C19, P450 2D6, and the wild type and allelic variants of P450 2C9, were prepared using similar methods. P450 1B1 is described by Shimada *et al.*, 1998. General schemes for the construction of plasmids for P450's 2A6, 2C19 and 2C9 are shown in Figures 1A, B and C, respectively.

Expression was done using TB media fortified with IPTG and δ-aminolevulinic acid (Gillam *et al.*, 1993; Gillam *et al.*, 1995). Bacterial membranes were prepared and incubated as described (Gillam *et al.*, 1993, Parikh *et al.*, 1997, Shimada *et al.*, 1998, Gillam *et al.*, 1997).

The nucleotide and corresponding amino acid sequences of P450 2A6, P450 2C19, P450 2C9 and P450 cin are provided in the sequence listing (see Table 1 for a summary) together with the associated proteins, hNPR and flavodoxin.

EXAMPLE 2

Observation of blue cultures

Blue pigment was seen in cultures co-expressing P450s with NPR. Initial TLC and visible spectroscopy with human, rat and mouse P450 2E1 systems yielded evidence for indigo production. The production of colour by P450 2A6 cultures was IPTG induction-dependent (Figure 2).

EXAMPLE 3

Isolation and characterization of pigments formed in bacterial cultures

A 1 litre culture of P450 2A6 was grown for 48 hr at 29°C. Centrifugation (at 200 x *g* for 1 min) yielded a dark blue pellet, which was collected and washed 5 times with H_2O (suspension and recentrifugation each time). The material was suspended in 10 ml of *N,N*-dimethylformamide and subjected to repeated sonication with a microprobe. $CHCl_3$ (100 ml) was added and the solution was washed five times with an equal volume of H_2O. The $CHCl_3$ layer was dried with Na_2SO_4 and filtered through paper (A_{600}=0.46, volume=156 ml). The material was concentrated at 50°C *in vacuo* to a low volume, which was streaked on a 1 mm x 20 cm x 20 cm silica gel G TLC plate. The plate was developed with $CHCl_3$-CH_3OH/50-1 v/v (Figure 3). Individual coloured bands were excised and extracted first with $CHCl_3$-CH_3OH 1-1 v/v and then with acetone, followed by centrifugation and concentration under N_2.

The extracellular pigment produced in a P450 2A6 culture was separated into four components, two blue and two pink/purple (Figure 3). Mass spectrometry yielded an apparent MH+ ion at *m/z* 263 for compounds 1, 2 and 4, indicating a molecular mass of 262. The visible spectra of components 2 and 4 (Figure 4, λ_{max} 603 nm and 552 nm, respectively) match the literature for indigo and indirubin, respectively (Laatsch and Ludwig-Kühn, 1986; Friedmann *et al.*, 1950; Fearon and Boggust, 1950), which both have molecular weights of 262. A definitive MH+ for compound 3 was not apparent, and the visible spectra did not match those of compound 2 or 4.

^1H NMR of compounds 2 and 4 provided further support for the assignment of these as indirubin and indigo (Hart *et al.*, 1992), respectively, particularly the 2-D spectra (Figures 5 and 6). The symmetric indigo structure is obvious, while indirubin shows the expected asymmetry.

UV-visible spectra were recorded in $CHCl_3$ using a Cary14/OLIS spectrophotometer (On-Line Instrument Systems, Bogart, GA). NMR spectra were recorded in d6-Me_2SO at 298 K

using Bruker AM-400 and AM-500 instruments (Bruker, Billerica, MA). Mass spectra were obtained in the positive ion mode using a Finnigan 7000 (Finnigan, Sunnyvale, CA) electrospray instrument, either by loop injection or by introduction following elution from a 2.1 x 150 mm Zorbax C8 column with a gradient of CH_3OH increasing from 4.5 to 81% v/v over 32 min in 0.1% w/v aqueous HCO_2H.

EXAMPLE 4

Assay of pigment production in vitro

Bacterial membranes containing recombinant human P450 (0.2 μM except P450 2D6 (0.05 μM)) and P450 1A1 (0.1 μM)) and NADPH-P450 reductase and human liver microsomes (0.2 μM P450) were prepared and incubated with indole (5 mM) under conditions supporting P450 catalysis (100 mM Tris-HCl, pH 7.4; 1 mM NADPH, 2.5mM glucose-6-phosphate, 0.5 U/ml glucose-6-phosphate dehydrogenase). The P450s most efficient in the production of the blue pigment co-migrating with indigo on TLC were 2A6>2C19~2E1 (Figure 7). However, lower amounts of other pigments were seen with most forms.

In order to estimate the relative production of pigments by different recombinant P450 forms expressed in bicistronic format with hNPR, bacterial membranes containing recombinant P450 enzymes and human NADPH-P450 reductase (0.2 μM P450) and human liver microsomes (0.04 μM P450), were incubated with 0.8 mM indole in the presence of an NADPH generating system as described above. Three preparations of bacterial membranes containing recombinant P450 2C19 and hNPR were tested corresponding to different levels of apparent pigment production during bacterial culture (1>2, 3). Bacterial membranes from cells transformed with the empty pCW vector or expressing recombinant hNPR alone were also included as minus-enzyme controls at protein concentrations equivalent to the mean of the protein concentrations in recombinant P450 incubations. Reactions were stopped at 0 (t=0 controls) or 30 mins by the addition of an equal volume of 1% w/v SDS and absorbance was measured at 665 nm (shown in preliminary studies to be the maximal wavelength for pigments produced by recombinant P450 2A6 when dissolved in the 0.5% w/v SDS milieu of quenched incubations). Figuyre

8 shows that significant pigment production was seen with a range of recombinant P450 preparations.

EXAMPLE 5

Further analysis of products of indole metabolism

In order to characterize the full range of products formed from indole, bacterial membranes containing recombinant P450 enzymes and human NADPH-P450 reductase, and rat and human liver microsomes (0.2 μM P450), were incubated with 5 mM indole under conditions supporting P450 catalysis (100 mM Tris-HCl, pH 7.4; 1 mM NADPH, 2.5 mM glucose-6-phosphate, 0.5 U/ml glucose-6-phosphate dehydrogenase). Products formed were extracted with methyl-t-butyl ether, desiccated and resuspended in 15% v/v acetonitrile: 85% v/v 50 mM potassium phosphate, pH 7.4, then subjected to high performance liquid chromatography using a 150 x 3.9 mm Waters C8 Symmetry reverse phase column and eluted at 1 ml/min with the following gradient: 0-15 mins 15% v/v acetonitrile, 85% v/v 50 mM potassium phosphate, pH 7.4; 15-30 mins 15% to 30% v/v acetonitrile gradient; 30-35 mins 30% v/v to 15% v/v acetonitrile; 35 to 45 mins isocratic at 15% v/v acetonitrile. Under these conditions, isatin eluted at 7 mins and indole at 35 mins. Other peaks, which were later identified by mass spectrometry as dioxindole and oxindole, eluted at 4.05 and 10.13 mins, respectively. Additional metabolite peaks eluting at times indicated in Table 3 below were not identified but relative yields were compared among incubations containing different P450 enzymes.

TABLE 3

Comparison of metabolite production from indole in incubations with various cytochrome p450 preparations

Peak retention time	Assignment	Relative metabolite formation *in vitro* (arbitrary units of peak area)[a]											
		hNPR only	2C19	1A2	1A1	2C9	2A6	2D6	2E1	1B1	3A4	3A5	HL
3.66		0.0	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4.05	**dioxindole**	4.3	9.1	1.7	2.9	13.6	9.2	6.5	7.3	7.7	6.9	5.0	0.2
4.52		0.0	2.5	0.9	0.0	0.5	0.8	0.0	1.2	0.8	0.5	0.0	0.0
5.63		0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0
7.37	**isatin**	5.9	43.5	14.9	3.5	23.1	47.9	19.8	44.0	17.2	10.9	5.7	13.9
8.03		1.0	6.2	7.2	1.2	1.6	1.0	1.2	0.8	0.9	0.6	1.2	2.1
8.9		0.0	8.5	2.5	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
10.13	**oxindole**	2.4	46.5	6.8	5.1	28.9	52.1	16.8	37.8	6.9	8.1	5.4	14.9
12.56		0.0	2.1	1.2	0.0	0.9	1.4	0.9	1.4	4.8	2.3	1.9	0.6
15.77		0.0	3.6	2.7	0.4	1.5	1.0	1.4	2.2	0.6	0.2	0.7	1.7
25.89		0.4	6.1	0.0	0.0	2.8	10.8	0.9	3.8	2.8	0.9	0.9	0.6
26.71		0.3	1.3	0.0	0.0	0.8	0.0	0.0	0.0	2.4	0.5	0.4	0.0
30.23		5.4	68.9	19.2	5.9	16.8	27.9	9.2	45.6	24.9	19.2	10.5	8.0

Peak retention time	Assignment	Relative metabolite formation *in vitro* (arbitrary units of peak area)[a]											
31.27		0.0	4.3	4.3	0.0	1.3	0.0	0.7	2.1	0.0	0.4	0.0	15.5
36.23		0.0	9.7	5.9	0.0	2.2	2.8	2.2	8.9	1.7	0.7	0.4	1.3
38.5		0.0	6.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
39.25		3.4	15.2	5.1	5.0	6.8	11.9	3.9	17.6	7.3	10.5	5.8	3.4
40.4		2.1	15.5	3.6	4.6	7.4	8.9	2.2	11.0	6.5	5.0	6.3	20.5

[a] Relative metabolite production was compared by measuring integrated peak areas in HPLC chromatograms. Incubations contained the P450 forms indicated in bacterial membranes with co-expressed hNPR. "hNPR only" indicates incubations using bacterial membranes containing hNPR but no recombinant P450 (negative control). "HL" denotes incubations containing microsomes from a single human liver rather than recombinant enzymes. Results are the mean of two determinations.

EXAMPLE 6

Characterization of the metabolic pathway:identification of products generated in vitro by recombinant bacterial membranes containing P450 2A6 and hNPR and by human liver microsomes

In order to define the primary and secondary metabolic pathways catalyzed by human P450 forms, indole (5 mM) was incubated with *E. coli* membranes containing P450 2A6 and hNPR for varying time periods (5-30 min) and the products were separated by HPLC using slight modifications (lower CH_3CN concentrations, larger size column - 10 x 250 mm, C_{18} instead of C_8) of the methods described in Example 5. UV spectra were measured on-line using a Thermo-Separations UV6000 rapid scanning spectrophotometer (Thermo-Separations, Piscataway, NJ, USA). Under these conditions, the substrate indole was eluted at 29.4 min (identified by co-chromatography with an external standard and mass spectrum, MH+ at *m/z* 118). Five of the product peaks eluted earlier and were identified by their spectral properties:

Dioxindole: t_R 7.5 min, UV λ_{max} 210 nm, 254 nm, 295 nm and spectrum identical to that published by Ward (1923) and Conforth *et al.* (1951). The product of $Na_2S_2O_4$ reduction of isatin (Sumpter, 1945) yielded the same compound as judged by HPLC t_R and UV spectrum.

Isatin: t_R 12.6 min, UV λ_{max} 244 nm, 305 nm. t_R and spectrum identical to commercial product (Aldrich) and products reported by Julian and Printy (1953) and Ward (1923). Mass spectrometry: *m/z* 148 (MH+).

Indoxyl: t_R 14.7 min, UV λ_{max} 239 nm. t_R and UV spectra identical to material generated by reaction of commercial indoxy-3-acetate and hog liver esterase (Sigma). Mass spectrometry: *m/z* 134 (MH+).

Oxindole: t_R 18.2 min, UV λ_{max} 203 nm, 248 nm, both the same as commercial material (Cornforth *et al.*, 1951; Barth *et al.*, 1972). Mass spectrometry: *m/z* 134 (MH+). NMR

($CDCl_3$) δ 3.56 (s, 2H, H-1), 6.88 (d, 1H, H-4 or H-7), 7.04(t, 1H, H-5 or H-6), 7.24(t, 1H, H-5 or H-6), 7.27 (d, 1H, H-4 or H-7), 7.53 (bs, $_1$H, NH).

6-Hydroxyindole: t_R 20.3 min, UV λ_{max} 218 nm, 293 nm. Spectrum identical to literature (Armerego, 1971; Stoll *et al.*, 1955). Mass spectrometry: *m/z* 134 (MH+).

No peaks produced by recombinant P450 2A6 or human liver microsomes corresponded to 4- or 5-hydroxyindole (commercial samples; Sigma), as judged by t_R and UV spectra.

The time course for production of the major products from indole is presented in Figure 9. At each indicated time point, a reaction aliquot was withdrawn and extracted into CH_2Cl_2 to stop the reaction. Products (CH_2Cl_2 phase) were concentrated and analyzed by HPLC, with the amount of product expressed on the basis of A_{220} units. Indoxyl was detected only in the early portion of the experiment. The production of indigo, as estimated by the A_{615} in the combined CH_2Cl_2 extract and a fraction prepared from the residual aqueous phase by adding two volumes of dimethylformamide, showed a lag but began within 2 min. The time courses of indoxyl and indigo appear to be related.

Oxindole appears to be the major primary oxidation product of indole formed by P450 2A6. The concentration of indole was varied in 2 min incubations with membranes containing P450 2A6 and hNPR. The parameters k_{cat} = 6.9 (±0.6) min^{-1} and Km 122 (±30) μM were determined.

When isatin was used as the substrate in similar incubations, it was rapidly reduced to dioxindole in the P450 2A6 system (only product). Subsequent experiments with purified hNPR showed that this enzyme catalyzed the reduction efficiently itself and the P450 2A6 did not seem to play a major role. The rate of reduction measured(with 1 mM isatin) was 290 min^{-1}.

Oxindole (1 mM) was stable when added to the typical P450 2A6/hNPR membrane system; no significant product formation was observed.

These results support the scheme of indole metabolism in P450 2A6/hNPR/NADPH systems and in NADPH-fortified human and rat liver microsomes follows the general features shown in Figure 10.

EXAMPLE 7

Bacterial P450 monooxygenase

Similar experiments as those described above were conducted using a cytochrome P450 monooxygenase from *Citrobacter brakii,* designated P450 cin. The nucleotide and corresponding amino acid sequences are shown in <400>20 and <400>21, respectively. Bicistronic constructs were prepared with the associated protein being flavodoxin (<400>22 and <400>23). Again, bacterial culture medium turned blue and data indicated the presence of indigo.

EXAMPLE 8

Characterization of the potential of P450 forms to catalyze indoxyl oxidation

Whereas the preceding experiments had provided evidence that various P450 forms could catalyse the monooxygenation of indole to indoxyl and that this could lead to the formation of indigo and related compounds, it was of interest to determine whether multiple steps along the pathway leading to indigo could be catalysed by the cytochrome P450 system. In previous systems characterized in regard to indigo generation, the oxidation of indoxyl was considered to occur spontaneously in aerobic conditions. However, the experiments outlined above were undertaken in a reducing environment (e.g. 1 mM NADPH in the *in vitro* experiments), raising the possibility that spontaneous oxidation may not be the sole mechanism by which two indoxyl monomers were transformed into one molecule of indigo. The object of the following experiments was to determine whether the P450 enzyme system could catalyse steps in the pathway leading to indigo subsequent to the initial oxidation of indole.

Bacterial membranes containing 0.1 μM P450 recombinant P450 2E1 and human NADPH-P450 reductase were incubated with an indoxyl generating system (0.5 mM indoxyl acetate and 50 Units porcine liver esterase [Sigma Chemical Company, Castle Hill, Australia] under conditions supporting P450 catalysis (100 mM Tris-HCl, pH 7.4; 1 mM NADPH, 2.5 mM glucose-6-phosphate, 0.5 U/ml glucose-6-phosphate dehydrogenase). Reactions were started by the addition of the NADPH generating system and esterase and incubated at 37°C in a thermostatically controlled cell compartment of a Beckman DU640 UV-Vis spectrophotometer. Indigo production was monitored by absorbance at 602 nm.

Two different preparations of membranes containing recombinant P450 2E1 and hNPR were tested, differing in the ratio of P450 to reductase (1:1.5 in preparation 1 and 1:0.95 in Preparation 2 in Table 4 below). Three different negative controls were used: incubations carried out in the absence of bacterial membranes; incubations carried out with bacterial membranes containing neither P450 nor hNPR, matched for protein concentration to the full incubations; and incubations containing bacterial membranes containing only hNPR (no P450), matched to the hNPR concentration in full incubations. Table 4 shows that incubations containing P450 2E1 showed elevated indigo production compared to all three controls. While the mean rate of indigo production in the first negative control (no bacterial membranes) was constant between the two sets of assays and rates for the other negative controls were similar. Incubations containing Preparation 2 showed a markedly higher rate of catalysis of indigo formation compared to Preparation 1, a difference that may be due to the different P450 2E1:hNPR ratios in the two preparations.

TABLE 4

Indigo production in incubation with recombinant P450 2E1 using an indoxyl generating system

Incubation	Rate of Indigo production (Absorbance units/min)[a]	Fold increase over highest control
P450 2E1/hNPR – Preparation 1	0.028 +/-0.002	1.8
No membranes	0.006 +/-0.001	
Bacterial membranes containing no P450 2E1 or hNPR	0.013 +/-0.002	
Bacterial membranes containing hNPR only	0.016 +/-0.001	
P450 2E1/hNPR – Preparation 2	0.040 +/-0.007	3.1
No membranes	0.006 +/-0.001	
Bacterial membranes containing no P450 2E1 or hNPR	0.013 +/-0.001	
Bacterial membranes containing hNPR only	0.013 +/-0.001	

[a] Values represent the mean +/- standard deviation of three independent measurements.

These results support the scheme of indigo and indirubin production catalyzed by P450 2A6/hNPR/NADPH systems in in NADPH-fortified human and rat liver microsomes as shown in Figure 11.

EXAMPLE 9

Construction of vectors for transient expression of recombinant human cytochrome P450 and reductase in plant tissue

Plant vectors containing the recombinant human cDNAs were prepared by subcloning the cDNAs for cytochrome P450 forms and hNPR into the vectors pGTVa (containing the 35S promoter) and p2.5GEM-11 (containing the AIM-1promoter) using the strategies outlined in Figure 12. Briefly, cDNA sequences for human cytochrome P450 forms and hNPR, which had been modified at the 5' end of the coding sequences to enhance expression in bacteria, were further modified to enhance expression in the eukaryotic (plant) system. This was effected by PCR-based oligonucleotide-directed mutagenesis of a short segment of the 5' coding sequence of the cDNA. Mutagenic primers were also designed to incorporate a restriction site to enable subcloning into the cognate vector and a ribosomal binding site sequence upstream of the translation start site (initial ATG codon). The resultant PCR-amplified DNA fragments were digested with appropriate restriction endonucleases and ligated into pGTVa or p2.5GEM-11 together with fragments encoding the remainder of the respective human cDNAs.

Table 5 shows alignments of approximately the first thirty nucleotides encoding cytochrome P450 2AE6, P450 2C19 and P450 2E1 monooxygenases, which highlight the differences required for expression in bacterial relative to the native eukaryotic sequence for use in plant cells.

TABLE 5

5' nucleotide sequence alignments of P450 and hNPR cDNAs: Comparison of sequences used for expression of bacterial and plant cells compared to native sequences

P450 2A6 cDNA

native ATGCTGGCCTCAGGGATGCTTCTGGTGGCCTTGCTGGTCTGCCTGACTGTGATGGTCTTGATGTCTGTTTGGCAGCAGAGGAAGAGCAAG
E.coli atggctgcttcaggaatgttattagtggccttgctggtctgcctgactgtgatggtcttgatgtctgtttggcagcagaggaagagcaag
plant atgctggcctcagggatgcttctggtggccttgctggtctgcctgactgtgatggtcttgatgtctgtttggcagcagaggaagagcaag

P450 2C19 cDNA

native ATGGATCCTTTTGTGGTCCTTGTGCTCTGTCTCTCATGTTTGCTTCTCCTTTCAATCTGGAGACAGAGCTCTGGGAGAGGAAAACTCCCTC
E.coli atggctcgacaatcttctggacgaggaaaactccctc
plant atggctagacagagctctgggagaggaaaactccctc

continued:

native CTGGCCCCACTCCTCTCCCAGTGATTGGAAATATCCTACAGATAGATATTAA
E.coli ctggccctactcctctcccagtgattggaaatatcctacagatagatattaa
plant ctggccctactcctctcccagtgattggaaatatcctacagatagatattaa

P450 2E1 cDNA

native ATGTCTGCCCTCGGAGTGACCGTGGCCCTGCTGGTGTGGGCGGCCTTCCTCCTGCTGGTGTCCATGTGGAGGCAGGTGCACAGCAGCTGG
E.coli atggctcgtcaagttcattcttcttggaatctgcccccaggccctttcccgcttcccatcatcgggaacctcttccagttggaattgaag
plant atgtctgccctcggagtCaccgtggccctgctggtgtgggcggccttcctcctgctggtgtccatgtggaggcaggtgcacagcagctgg

hNPR cDNA

native atgggagactcccacgtggacaccagctccaccgtgtccgaggcggtggccgaagaagtatctcttttcagcatgacggacatgattctg
E.coli atggctgactcccacgtggacaccagctccaccgtgtccgaggcggtggccgaagaagtatctcttttcagcatgacggacatgattctg
plant atgggagactcccacgtggacaccagctccaccgtgtccgaggcggtggccgaagaagtatctcttttcagcatgacggacatgattctg

EXAMPLE 10

Bombardment of plant tissue with genetic material encoding a human P450 enzyme and/or human reductase to effect transient expression of a functional recombinant cytochrome P450 system

The aim of these experiments is to transiently express a human P450 enzyme with or without human NADPH-P450 reductase (hNPR) in a white flower petal, namely petunia and to determine whether expression could be detected by the visible production of a blue pigment in plant tissue, namely indigo. A secondary aim is to determine whether indigo production might be enhanced by supplementation with low concentrations of indole.

The gene bombardment protocol is initially optimised using the reporter vector pGTVa-GUS. GUS expression is assayed using the method described by Jefferson *et al.*, (1992). Efficiency of the transformation is measured by the mean number of blue spots per petal bombardment. The parameters examined during these initial optimisation experiments are target distance, bombardment pressure and petal developmental stage.

Plasmid DNA is obtained from *E.coli* using a standard alkaline lysis procedure with and without additional procedures for purification of the resultant DNA (Sambrook *et al.*, 1989). The DNA is prepared for bombardment by combining various amounts of tungsten particle solution with DNA. After vortexing, the particles are precipitated with $CaCl_2$ and spermidine. After removing a portion of the supernatant, the tungsten suspension was vortexed and an aliquot removed for bombardment.

In this experiment, white petunia flowers are used for bombardment. Petunia plants having other colours may also be used. The device used for bombardment is the particle inflow gun developed by Finer *et al.*, (1992) which propels tungsten particles directly in a stream of helium towards the target. The petal is placed in a petri dish containing filterpaper moistened with either sterile water or a solution of indole or osmotic medium (with or without indole). Each petal preparation is bombarded with one of:-

(a) vector containing a P450 cDNA alone; or

(b) vector containing a P450 cDNA plus vector containing the hNPR cDNA; or

(c) vector containing the GUS control cDNA.

In some cases the vector containing the GUS control cDNA is bombarded simultaneously with either or both types of vector containing human-derived sequences.

The petal distance and helium pressure are as found optimal in the initial optimization experiments. A negative control containing tungsten particles only is included for both the water and indole petal samples.

The success of the bombardment is analyzed by the presence of blue spots after overnight incubation of the bombarded petal in either water or indole solution or osmotic medium (indicating *in situ* production of indigo).

EXAMPLE 11

Introduction of genetic mutant encoding P450 monooxygenase to cells of non-white plant

Different shades of colours are proposed to be obtainable using different coloured recipient plants for the genetic material encoding the cytochrome P450 monooxygenase/associated protein of the present invention. Accordingly, plants having coloured flowers such as red, blue, yellow, orange, purple, pink, spotted and/or mottled are selected as recipients of the subject genetic material. Various colour combinations are expected which greatly enhances the variety of colour phenotypes resulting from the use of the instant genetic constructs.

EXAMPLE 12

Transformation procedures

Yeast strains are transformed with genetic material according to Ito *et al.*, (1990). Genetic material is introduced into the *Agrobacterium tumefaciens* strain AGL0 by adding 5 μg of plasmid DNA to 100 μl of competent AGL0 cells prepared by inoculating a 50 ml culture

of MG/L (Garfinkel and Nester, 1980). These are cultured and grown for 16 hours with shaking at 28°C. The cells are then pelleted and resuspended in 0.5 ml of 85% v/v 100 mM $CaCl_2$/15% v/v glycerol. The DNA-*Agrobacterium* mixture is frozen by incubation in liquid N_2 for 2 minutes and then allowed to thaw by incubation at 37°C for 5 minutes. The DNA/bacterial mix is then placed on ice for a further 10 minutes. The cells are then mixed with 1 ml of LB (Sambrook *et al.*, 1989) media and incubated with shaking for 16 hours at 28°C. Cells of *A. tumefaciens* carrying genetic material are selected on LB agar plates containing 10 μg/ml gentamycin or other suitable selection such as another antibiotic or a herbicide. The presence of genetic material is confirmed by Southern analysis of DNA isolated from the gentamycin-resistant transformants or any other selectable molecule such as another antibiotic or a herbicide.

Petunia transformations

(a) Plant Material

Leaf tissue from mature plants is treated in 1.25% w/v sodium hypochlorite for 2 minutes and then rinsed three times in sterile water. The leaf tissue is then cut into 25 mm^2 squares and precultured on MS media (Murashige and Skoog, 1962) supplemented with 0.05 mg/l kinetin and 1.0 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D) for 24 hr.

(b) Co-cultivation of Agrobacterium Tissue

Agrobacterium tumefaciens strain AGL0 containing genetic material is maintained at 4°C on MG/L agar plates with 100 mg/L gentamycin. A single colony is grown overnight in liquid medium containing 1% w/v Bacto-peptone, 0.5% w/v Bacto-yeast extract and 1% w/v NaCl. final concentration of 5 x 10^8 cells/ml is prepared the next day by dilution in liquid MS medium containing B5 vitamins (Gamborg *et al,* 1968) and 3% w/v sucrose (BPM). The leaf discs were dipped for 2 minutes into BPM containing AGL0/genetic material. The leaf discs are then blotted dry and placed on co-cultivation media for 4 days. The co-cultivation medium consists of SH medium (Schenk and Hildebrandt, 1972) supplemented with 0.05 mg/l kinetin and 1.0 mg/l 2,4-D and included a feeder layer of

tobacco cell suspension spread over the co-cultivation medium with a filter paper placed on top of the tobacco cell suspension.

(c) Recovery of transgenic plants

After co-cultivation, the leaf discs are transferred to a selection medium (MS medium supplemented with 3% w/v sucrose, α-benzylaminopurine (BAP) 2 mg/l, 0.5 mg/l α-naphthalene acetic acid (NAA), kanamycin 300 mg/l, 350 mg/L cefotaxime and 0.3% w/v Gelrite Gellan Gum (Schweizerhall)). Regenerating explants are transferred to fresh selection medium after 4 weeks. Adventitious shoots which survive the kanamycin selection are isolated and transferred to BPM containing 100 mg/l kanamycin and 200 mg/l cefotaxime for root induction. All cultures are maintained under a 16 hour photoperiod (60 μmol. m^{-2}, s^{-1} cool white fluorescent light) at 23 ± 2°C. When roots reach 2-3 cm in length the transgenic petunia plantlets are transferred to autoclaved Debco 51410/2 potting mix in 8 cm tubes. After 4 weeks, plants are replanted into 15 cm pots, using the same potting mix, and maintained at 23°C under a 14 hour photoperiod (300 μmol. m^{-2}, s^{-1} mercury halide light).

EXAMPLE 13

Transformation of Dianthus caryophyllus

(a) Plant material

Dianthus caryophyllus, (cv. Crowley Sim, Red Sim, Laguna) cuttings are used in this experiment. The outer leaves are removed and the cuttings are sterilized briefly in 70% v/v ethanol followed by 1.25% w/v sodium hypochlorite (with Tween 20) for 6 minutes and rinsed three times with sterile water. All the visible leaves and axillary buds are removed under the dissecting microscope before co-cultivation.

(b) Co-cultivation of Agrobacterium and Dianthus Tissue

Agrobacterium tumefaciens strain AGL0 containing a genetic construct encoding a cytochrome P450 monooxygenase and optionally an associated protein as herein described is maintained at 4 on MG/l (Garfinkel and Nester, 1980) agar plates with 100 mg/l gentamycin. A single colony is grown overnight in liquid MG/L broth and diluted to 5 x 10^8 cells/ml the next day before inoculation. *Dianthus* tissue is co-cultivated with *Agrobacterium* on MS medium (Murashige and Skoog, 1962) supplemented with 3% w/v sucrose, 5 mg/l α-naphthalene acetic acid (NAA), 20 μM acetosyringone and 0.8% w/v Difco Bacto Agar (pH 5.7).

(c) Recovery of Transgenic Dianthus Plants

Co-cultivated tissue is transferred to MS medium supplemented with 1 mg/l benzylaminopurine (BAP), 0.1 mg/l NAA, 150 mg/l kanamycin, 500 mg/l ticarcillin and 0.8% w/v Difco Bacto Agar (selection medium). After three weeks, explants are transferred to fresh selection medium and care is taken at this stage to remove axillary shoots from stem explants. After 6-8 weeks on selection medium, healthy adventitious shoots are transferred to hormone free MS medium containing 3% w/v sucrose, 150 mg/l kanamycin, 500 mg/l ticarcillin, 0.8% w/v Difco Bacto Agar. At this stage, GUS histochemical assay (Jefferson, 1987) and/or NPT II dot-blot assay (McDonnell *et al.*, 1987) are used to identify transgenic shoots. Transgenic shoots are transferred to MS medium supplemented with 3% w/v sucrose, 500 mg/l ticarcillin and 0.4% w/v Gelrite Gellan Gum (Schweizerhall) for root induction. All cultures are maintained under a 16 hour photoperiod (120 μE cool white fluorescent light) at 23 ± 2°C. When plants are rooted and reached 4-6 cm tall they are acclimatised under mist. A mix containing a high ratio of perlite (75% or greater) soaked in hydroponic mix (Kandreck and Black, 1984) is used for acclimation, which typically lasts 4-5 weeks. Plants are acclimatised at 23°C under a 14 hour photoperiod (200 μE mercury halide light).

EXAMPLE 14

Transformation of Rosa hybrida

1. ***Rosa hybrida* cv Royalty**

Plant tissues of the rose cultivar Royalty are transformed according to the method disclosed in PCT/AU91/04412, having publication number WO92/00371.

2. ***Rosa hybrida* cv Kardinal**

(a) Plant Material

Kardinal shoots are used. Leaves are removed and the remaining shoots (5-6 cm) are sterilized in 1.25 % w/v sodium hypochlorite (with Tween 20) for 5 minutes followed by three rinses with sterile water. Isolated shoot tips are soaked in sterile water for 1 hr and precultured for 2 days on MS medium containing 3% w/v sucrose, 0.1 mg/L BAP, 0.1 mg/l kinetin, 0.2 mg/l Gibberellic acid, 0.5% w/v polyvinyl pyrrolidone and 0.25% w/v Gelrite Gellan Gum, before co-cultivation.

(b) Co-cultivation of Agrobacterium and Rosa shoot Tissue

Agrobacterium tumefaciens strains ICMP 8317 (Janssen and Gardner, 1989) and AGL0, containing genetic constructs encoding a cytochrome P450 monooxygenase and optionally an NPR are maintained at 4°C on MG/L agar plates with 100 mg/L gentamycin. A single colony from each *Agrobacterium* strain is grown overnight in liquid MG/L broth. A final concentration of 5 x 10^8 cells/ml is prepared the next day by dilution in liquid MG/L. Before inoculation, the two *Agrobacterium* cultures are mixed in a ratio of 10:1. A longitudinal cut is made through the shoot tip and an aliquot of 2 µl of the mixed *Agrobacterium* cultures is placed as a drop on the shoot tip. The shoot tips are co-cultivated for 5 days on the same medium used for preculture.

Agrobacterium tumefaciens strain AGL0 is maintained at 4°C on MG/L agar plates with 100 mg/L kanamycin. A single colony from each *Agrobacterium* strain is grown overnight

in liquid MG/L broth. A final concentration of 5 x 10^8 cells/ml is prepared the next day by dilution in liquid MG/L.

(c) Recovery of Transgenic Rosa Plants

After co-cultivation, the shoot tips are transferred to selection medium. Shoot tips are transferred to fresh selection medium every 3-4 weeks. Galls observed on the shoot tips are excised when they reached 6-8 mm in diameter. Isolated galls are transferred to MS medium containing 3% w/v sucrose, 25 mg/l kanamycin, 250 mg/l cefotaxime and 0.25% w/v Gelrite Gellan Gum for shoot formation. Shoots regenerated from gall tissue are isolated and transferred to selection medium. GUS histochemical assay and callus assay are used to identify transgenic shoots. Transgenic shoots are transferred to MS medium containing 3% w/v sucrose, 200 mg/l cefotaxime and 0.25% w/v Gelrite Gellan Gum for root induction. All cultures are maintained under 16 hour photoperiod (60 µE cool white fluorescent light) at 23 ± 2°C. When the root system is well developed and the shoot reached 5-7 cm in length the transgenic rose plants are transferred to autoclaved Debco 514110/2 potting mix in 8 cm tubes. After 2-3 weeks plants are replanted into 15 cm pots using the same potting mix and maintained at 23°C under a 14 hour photoperiod (300 µE mercury halide light). After 1-2 weeks potted plants are moved to glasshouse (Day/Night temperature : 25-28°C/14°C) and grown to flowering.

EXAMPLE 15

Transformation of *Chrysanthemum morifolium*

(a) Plant Material

Chrysanthemum morifolium (cv. Blue Ridge, Pennine Chorus) cuttings are obtained. Leaves are removed from the cuttings, which were then sterilized briefly in 70% v/v ethanol followed by 1.25% w/v sodium hypochlorite (with Tween 20) for 3 minutes and rinsed three times with sterile water. Internodal stem sections are used for co-cultivation.

(b) Co-cultivation of Agrobacterium and Chrysanthemum Tissue

Agrobacterium tumefaciens strain LBA4404 (Hoekema *et al.*, 1983), containing is grown on MG/l agar plates containing 50 mg/l rifampicin and 10 mg/l gentamycin. A single colony from each *Agrobacterium* is grown overnight in the same liquid medium. These liquid cultures are made 10% v/v with glycerol and 1 ml aliquots transferred to the freezer (-80°C). A 100-200 µl aliquot of each frozen *Agrobacterium* is grown overnight in liquid MG/l containing 50 mg/l rifampicin and 10 mg/l gentamycin. A final concentration of 5 x 10^8 cells/ml is prepared the next day by dilution in liquid MS containing 3% w/v sucrose. Stem sections are co-cultivated with *Agrobacterium* in co-cultivation medium for 4 days.

(c) Recovery of Transgenic Chrysanthemum Plants

After co-cultivation, the stem sections were transferred to selection medium. After 3-4 weeks, regenerating explants are transferred to fresh medium. Adventitious shoots which survive the kanamycin selection are isolated and transferred to MS medium containing kanamycin and cefotaxime for shoot elongation and root induction. All cultures are maintained under a 16 hour photoperiod (80 µE cool white fluorescent light) at 23 ± 2°C. Leaf samples are collected from plants which rooted on kanamycin and Southern blot analysis is used to identify transgenic plants. When transgenic chrysanthemum plants reach 4-5 cm in length, they are transferred to autoclaved Debco 51410/2 potting mix in 8 cm tubes. After 2 weeks, plants are replanted into 15 cm pots using the same potting mix and maintained at 23°C under a 14 hour photoperiod (300 µE mercury halide light). After 2 weeks potted plants are moved to glasshouse (Day/Night temperature : 25-28°C/14°C) and grown to flowering.

EXAMPLE 16

Transgenic plant phenotype analysis

The analysis of transgenic plant phenotypes in relation to changes in flower or plant tissue colour is dependent on the recipient plants. For example, where genetic material is introduced into plants having white flowers, plant flower colour is determined using colour

codes taken from the Royal Horticultural Society's Colour Chart (RHSCC). This provides one means by which to describe the colour phenotypes observed. The determination of designated numbers, however, should be taken only as a guide to the perceived colours and should not be regarded as limiting the possible colours which may be obtained.

Where a recipient plant comprises non-white flowers, a colour change may be assessed *inter alia* with respect to colour *per se,* uniformity of colour, colour distribution as well as chemically such as by analysing possible combinations of indigo or its derivatives or precursors and endogenous pigment molecules (e.g. flavonoid compounds).

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

BIBLIOGRAPHY

Armerego. Ultraviolet spectra of heterocycles. In: *Physical Methods in Heterocyclic Chemistry*, edited by Katritzky, A.R., Academic Press, New York, p. 95, 1971.

Barth *et al., Helvetica Chimica Acta, 55:* 2168-2178, 1972.

Bonner and Laskey. *European Journal of Biochemistry 46:* 83, 1974.

Cornforth,*et al., Biochemical Journal, 48:* 598-603, 1951.

De Vetten *et al., Proceedings of the National Academy of Science USA 96:* 778-783, 1999.

Fearon and Boggust, *Biochemical Journal 46:* 62-67, 1950

Finer *et al., Plant Cell Reports 11:* 323-328, 1992.

Friedmann *et al., Biochimica et Biophysica Acta 5:* 45-52.

Gamborg *et al, Experimental Cell Research 50:* 151-158, 1968.

Gerfinkel and Nester. *Journal of Bacteriology 144:* 732-743, 1980.

Gillam *et al., Archives of Biochemistry and Biophysics 305:* 123-131, 1993.

Gillam *et al., Archives of Biochemistry and Biophysics 319:* 540-550, 1995.

Gillam *et al., Archives of Biochemistry and Biophysics 346:* 81-90, 1997.

Hart *et al., Journal of General Microbiology 138:* 211-216, 1992.

Hoekema *et al., Nature 303:* 179-180, 1983.

Ito *et al., Journal of Bacteriology 153:* 163-168, 1983.

Jannsen and Gerdner, *Plant Molecular Biology 14:* 61-72, 1989.

Jefferson *et al., EMBO Journal 6(13):* 3901-3907, 1987.

Julian and Printy. *Journal of the American Chemical Society, 75:* 5301-5305, 1953.

Laatsch and Ludwig-Kühn, *Liebigs Annals of Chemistry* 1847-1853, 1986.

Lazo *et al., Biotechnology 9:* 963-967, 1991.

Marmur and Doty, *Journal of Molecular Biology 5:* 109, 1962.

McDonnell *et al., Plant Molecular Biology Reports 4:* 380-386, 1987.

Murashige and Skoog, *Plant Physiology 15:* 73-97, 1962.

Needleman and Wunsch, *Journal Molecular Biology 48:* 443-453, 1970.

Parikh *et al., Nature Biotechnology 15:* 784-788, 1997.

Sambrook *et al.,* (eds) *Molecular Cloning: A Laboratory Manual* (2nd ed). Cold Spring Harbor Laboratory Press, USA, 1989.

Schenk and Hilderbrandt, *Canadian Journal of Botany 50:* 199-204, 1972.

Shimada *et al., Archives of Biochemistry and Biophysics 357:* 111-120, 1998.

Stoll *et al., Helvetica Chimica Acta 38:* 1452-1472, 1955.

Sumpter, *Journal of the American Chemical Society 67:* 1140-1141, 1945.

Ward *et al., Biochemical Journal, 17:* 891-897, 1923.

CLAIMS

1. Use of genetic material encoding a cytochrome P450 protein or a functional derivative or homologue thereof in the generation of a genetically modified cell, which cell has the capacity to produce a pigment in the presence of indole or a precursor, analogue or derivative thereof upon expression of said genetic material.

2. Use according to Claim 1 wherein the genetic material, encoding the cytochrome P450 protein or its functional derivative or homologue, is co-expressed with genetic material encoding an associated protein.

3. Use according to Claim 2 wherein the genetic material encoding an associated protein is introduced into the cell.

4. Use according to Claim 2 wherein the genetic material encoding an associated protein is indigenous to the cell.

5. Use according to Claim 2 or 3 or 4 wherein the associated protein selected from a co-enzyme-dependent cytochrome P450 reductase and/or a cytochrome b_5 and/or a flavodoxin or a functional derivative or homologue of either molecule.

6. Use according to Claim 1 wherein the cytochrome P450 protein is of mammalian origin.

7. Use according to Claim 1 wherein the cytochrome P450 protein is of microbial origin.

8. Use according to Claim 6 wherein the cytochrome P450 protein is a co-enzyme-dependent cytochrome P450 monooxygenase.

9. Use according to Claim 8 wherein the cytochrome P450 protein is P450 2A6 encoded by a nucleotide sequence substantially as set forth in <400>17 or a nucleotide

sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>17 or its complementary form under low stringency conditions.

10. Use according to Claim 8 wherein the cytochrome P450 protein is P450 2E1 encoded by a nucleotide sequence substantially as set forth in <400>18 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>18 or its complementary form under low stringency conditions.

11. Use according to Claim 8 wherein the cytochrome P450 protein is P450 2C19 encoded by a nucleotide sequence substantially as set forth in <400>19 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>19 or its complementary form under low stringency conditions.

12. Use according to Claim 8 wherein the cytochrome P450 protein is P450 cin encoded by a nucleotide sequence substantially as set forth in <400>20 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

13. Use according to any one of the preceding claims wherein the genetically modified cell is of plant origin.

14. Use according to any one of Claims 1 to 12 wherein the genetically modified cell is of microbial origin.

15. A method of inducing a colour change in a plant or part of a plant, said method comprising introducing into plant cells a genetic construct encoding a cytochrome P450 protein or a functional derivative or homologue thereof, regenerating a plant from said plant cells and subjecting said plant to appropriate conditions sufficient to permit

expression of a nucleotide sequence on said genetic construct which encodes said cytochrome P450 protein such that in the presence of indole or a precursor, analogue or derivative thereof said cytochrome P450 protein facilitates the conversion of indole or its precursor, analogue or derivative to an intermediate which is capable of oxidation to a pigment.

16. A method according to Claim 15 wherein the genetic construct encoding the cytochrome P450 protein is co-expressed with genetic material encoding an associated protein.

17. A method according to Claim 16 wherein the associated protein is selected from a co-enzyme-dependent cytochrome P450 reductase and/or a cytochrome b_5 or a functional derivative or homologue of either molecule.

18. A method according to Claim 15 wherein the cytochrome P450 protein is of mammalian origin.

19. A method according to Claim 15 wherein the cytochrome P450 protein is of microbial origin.

20. A method according to Claim 15 wherein the cytochrome P450 protein is a co-enzyme-dependent cytochrome P450 monooxygenase.

21. A method according to Claim 20 wherein the cytochrome P450 protein is P450 2A6 encoded by a nucleotide sequence substantially as set forth in <400>17 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>17 or its complementary form under low stringency conditions.

22. A method according to Claim 20 wherein the cytochrome P450 protein is P450 2E1 encoded by a nucleotide sequence substantially as set forth in <400>18 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide

sequence capable of hybridizing to <400>18 or its complementary form under low stringency conditions.

23. A method according to Claim 20 wherein the cytochrome P450 protein is P450 2C19 encoded by a nucleotide sequence substantially as set forth in <400>19 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>19 or its complementary form under low stringency conditions.

24. A method according to Claim 20 wherein the cytochrome P450 protein is P450 cin encoded by a nucleotide sequence substantially as set forth in <400>20 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

25. A genetically modified cell or multicellular animal or plant or progeny thereof or parts of said transgenic plant or animal wherein said transgenic plant or animal comprises expressible genetic material encoding a microbial or mammalian cytochrome P450 protein.

26. A genetically modified cell or multicellular animal or plant or progeny or parts thereof according to Claim 25 wherein upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue or derivative is converted to an intermediate which is capable of oxidation to a pigment.

27. A genetically modified cell according to Claim 25 or 26 wherein the cell is a microbial cell.

28. A transgenic plant or progeny or parts thereof comprising expressible genetic material encoding a microbial or mammalian cytochrome P450 monooxygenase such that upon expression of said genetic material and in the presence of an associated protein and indole or a precursor, analogue or derivative thereof, the indole or its precursor, analogue

or derivative is converted to an intermediate which is capable of oxidation to a pigment.

29. A transfected or transformed cell, tissue, organ or non-cellular material which contains or is capable of producing cytochrome P450 protein or a functional derivative or homologue thereof.

30. A transfected or transformed cell, tissue, organ or non-cellular material according to Claim 29 wherein the cytochrome P450 protein is P450 2A6 encoded by a nucleotide sequence substantially as set forth in <400>17 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>17 or its complementary form under low stringency conditions.

31. A transfected or transformed cell, tissue, organ or non-cellular material according to Claim 29 wherein the cytochrome P450 protein is P450 2E1 encoded by a nucleotide sequence substantially as set forth in <400>18 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>18 or its complementary form under low stringency conditions.

32. A transfected or transformed cell, tissue, organ or non-cellular material according to Claim 29 wherein the cytochrome P450 protein is P450 2C19 encoded by a nucleotide sequence substantially as set forth in <400>19 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>19 or its complementary form under low stringency conditions.

33. A transfected or transformed cell, tissue, organ or non-cellular material according to Claim 29 wherein the cytochrome P450 protein is P450 cin encoded by a nucleotide sequence substantially as set forth in <400>20 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

34. Flowers, parts of plants, fibrous material from plants, non-cellular material and reproductive portions of plants which are coloured or have the capacity to be coloured in

the presence of indole or a precursor thereof due to the presence of cytochrome P450 protein and an associated protein.

35. Flowers, parts of plants, fibrous material from plants, non-cellular material and reproductive portions of plants according to Claim 34 wherein the colour produced in the presence of indole requires an introduced or indigenous associated protein.

36. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

37. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

38. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the

capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10 or <400>12;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4 or <400>6;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3 or <400>5; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

39. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an NPR wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(ii) a nucleotide sequence defined by <400>13 and <400>15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14 and <400>16;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13 and <400>15; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

40. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(ii) a nucleotide sequence defined by <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

41. Use of a genetic sequence comprising a nucleotide sequence expressible in plant cells which encodes a cytochrome P450 monooxygenase in the generation of a transgenic plant or transformed cells of a plant which, upon expression of said nucleotide sequence, and in the presence of indole or a precursor, analogue and/or derivative thereof, have the capacity to convert indole or a precursor, analogue or derivative thereof to an intermediate which is oxidizable to a pigment wherein the genetic sequence further comprises a nucleotide sequence encoding an bacterial flavodoxin wherein the nucleotide sequence is selected from:

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

42. Use of plant or plant material as a source of pigment for use, for example, in the dye or cosmetics industry or as commercial tags or as markers for transformation of microbial, plant or animal cells.

43. A genetic construct comprising a cistron encoding a microbial or mammalian cytochrome P450 protein for use in the generation of a transformed plant cell such that upon expression of said cistron, the cytochrome P450 facilitates conversion of indole or a precursor thereof to a molecule oxidizable to a pigment.

44. A genetic construct according to Claim 43 wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/P/W/L/Y)(G/S/C/N/H/E)X (G/D/A/E)X(R/H/S/K/T)XCX$_1$(G/A) wherein X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and wherein where residues in parentheses represent alternatives at a single position.

45. A genetic construct according to Claim 43 wherein the cytochrome P450 monooxygenase comprises the amino acid sequence (F/L)(G/S)XGX(R/H)XCX$_1$(G/A) where X is any amino acid and X$_1$ is selected from V, M, A, L, I, P, F or T and where residues in parentheses represent alternatives at a single position.

46. A genetic construct according to Claim 43 or 44 or 45 comprising a nucleotide sequence selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>8, <400>10, <400>12 or <400>21;

(ii) a nucleotide sequence defined by <400>7, <400>9 or <400>11 or <400>20;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>2, <400>4, <400>6 or <400>21;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>1, <400>3, <400>5 or <400>20; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

47. A genetic construct according to Claim 43 or 44 or 45 or 46 further comprising a nucleotide sequence encoding an associated protein human NPR (hNPR) wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>16;

(ii) a nucleotide sequence defined by <400> 15;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>14;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>13; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

48. A genetic construct according to Claim 43 or 44 or 45 or 46 further comprising a nucleotide sequence encoding bacterial flavodoxin wherein the nucleotide sequence is selected from:-

(i) a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(ii) a nucleotide sequence defined by <400>22;

(iii) a nucleotide sequence having at least 50% similarity to a nucleotide sequence in (i) or (ii);

(iv) a nucleotide sequence having at least 50% similarity to a nucleotide sequence encoding the amino acid sequence defined in <400>23;

(v) a nucleotide sequence having at least 50% similarity to a nucleotide sequence defined in <400>22; or

(vi) a nucleotide sequence capable of hybridizing under low stringency conditions to a nucleotide sequence in one or more of (i), (ii), (iii), (iv) or (v).

49. An isolated nucleic acid molecule comprising a nucleotide sequence encoding or complementary to a sequence encoding a cytochrome P450 cin monooxygenase from *Citrobacter brakii*.

50. An isolated nucleic acid molecule according to Claim 49 wherein the cytochrome P450 monooxygenase comprises an amino acid sequence substantially as set forth in <400>21 or an amino acid sequence having at least about 50% similarity thereto after optimal alignment.

51. An isolated nucleic acid molecule according to Claim 50 comprising a nucleotide sequence substantially as set forth in <400>20 or a nucleotide sequence having at least about 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

52. A method of inducing a colour change in or by a microbial cell, said method comprising introducing into microbial cells a genetic construct encoding a cytochrome P450 protein or a functional derivative or homologue thereof, and subjecting said cells to appropriate conditions sufficient to permit expression of a nucleotide sequence on said genetic construct which encodes siad cytochrome P450 protein such that in the presence of indole or a precursor, analogue or derivative thereof said cytochrome P450 protein facilitates the conversionof indole or its precursor, analogue or derivative to an intermediate which is capable of oxidation to a pigment.

53. A method according to Claim 52 wherein the genetic construct encoding the cytochrome P450 protein is co-expressed with genetic material encoding an associated protein.

54. A method according to Claim 53 wherein the associated protein is selected from a co-enzyme-dependent cytochrome P450 reductase and/or a cytochrome b_5 and/or a flavodoxin or a functional derivative or homologue of either molecule.

55. A method according to Claim 52 wherein the cytochrome P450 protein is of mammalian origin.

56. A method according to Claim 52 wherien the cytochrome P450 protein is of microbial origin.

57. A method according to Claim 55 wherein the cytochrome P450 protein is a co-enzyme-dependent cytochrome P450 monooxygenase.

58. A method according to Claim 57 wherein the cytochrome P450 protein is P450 2A6 encoded by a nucleotide sequence substantially as set forht in <400>17 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>17 or its complementary form under low stringency conditions.

59. A method according to Claim 57 wherein the cytochrome P450 protein is P450 2E1 encoded by a nucleotide sequence substantially as set forth in <400>18 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide seuqence capable of hybridizing to <400>18 or its complementary form under low stringency conditions.

60. A method according to Claim 57 wherein the cytochrome P450 protein is P450 2C19 encoded by a nucleotide sequence substantially as set forth in <400>19 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>19 or its complementary form under low stringency conditions.

61. A method according to Claim 57 wherein the cytochrome P450 protein is P450 cin encoded by a nucleotide sequence substantially as set forth in <400>20 or a nucleotide sequence having at least 50% similarity thereto after optimal alignment or a nucleotide sequence capable of hybridizing to <400>20 or its complementary form under low stringency conditions.

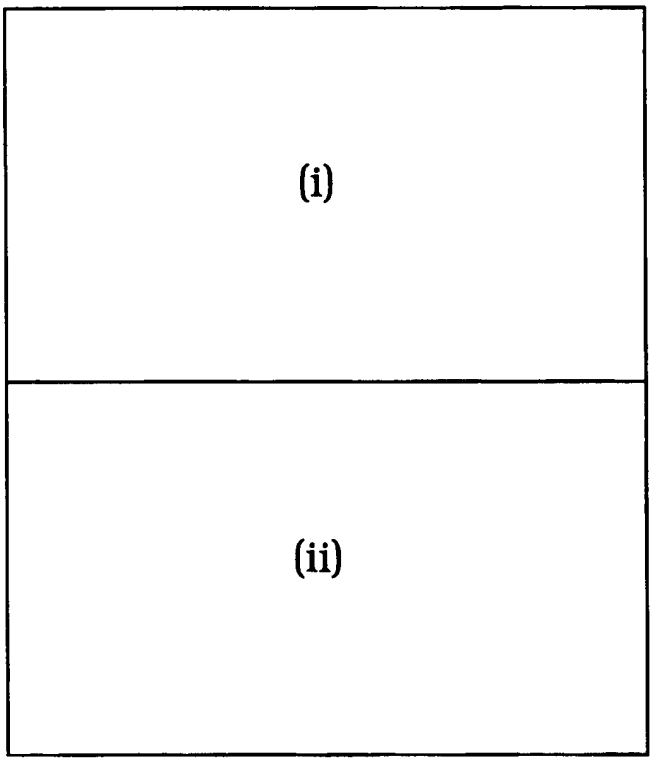

Figure 1a

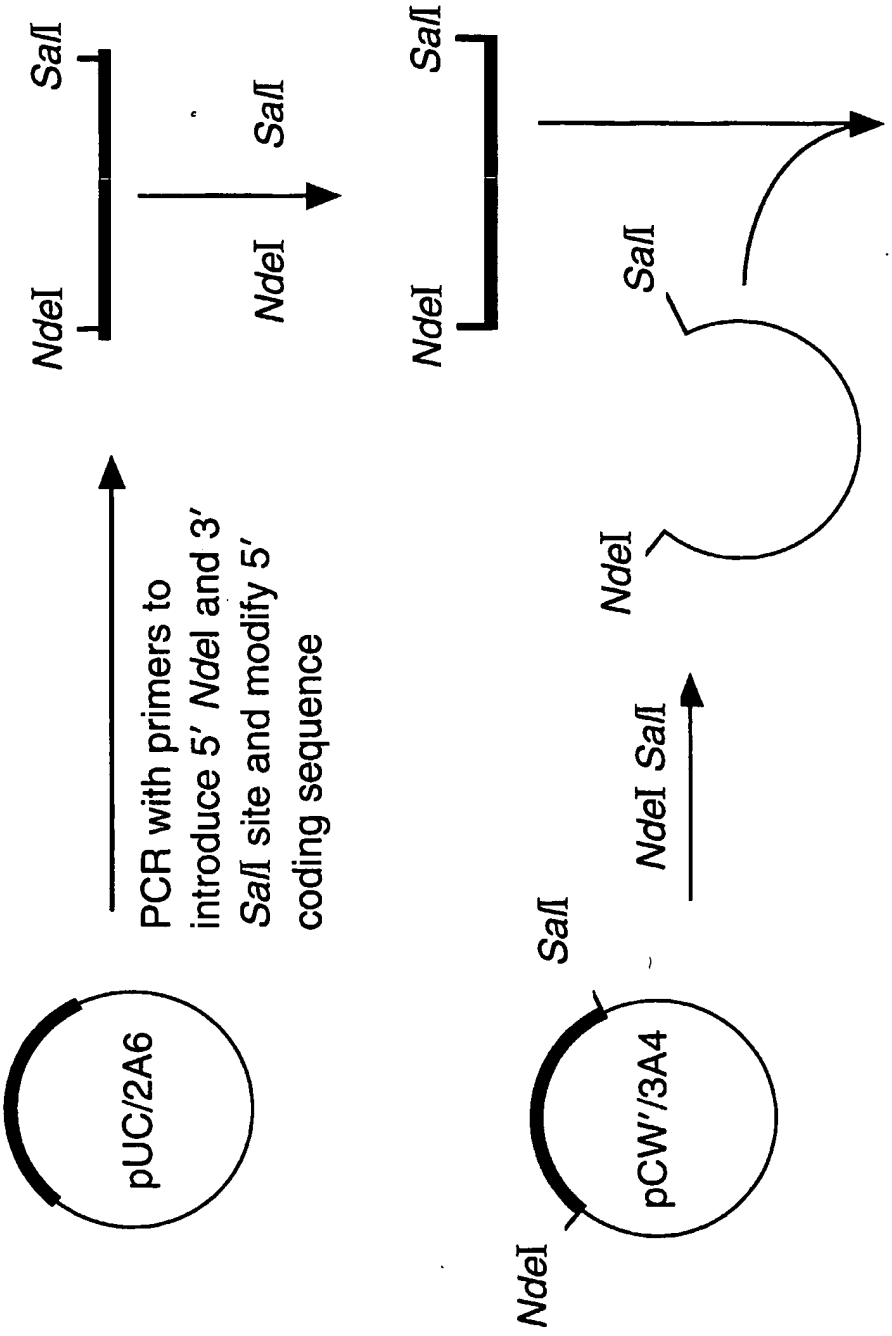
SalI
NdeI
PCR with primers to introduce 5' NdeI and 3' SalI site and modify 5' coding sequence
pUC/2A6
SalI
NdeI
pCW'/3A4
NdeI SalI
NdeI
SalI
NdeI
SalI

Figure 1a(i)

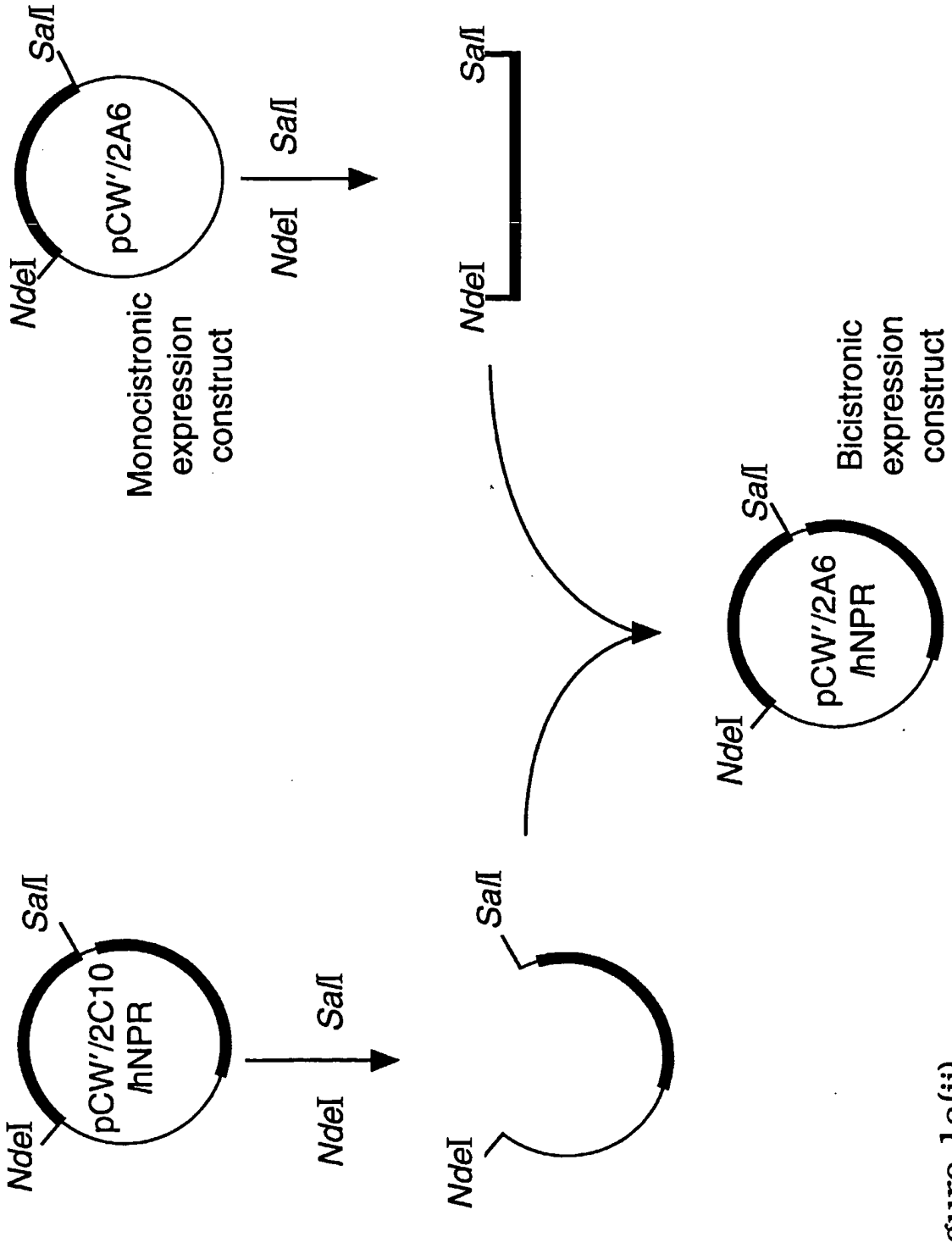
SalI
NdeI
pCW"/2A6
Monocistronic expression construct
SalI
NdeI
SalI
NdeI
pCW"/2C10 hNPR
SalI
NdeI
SalI
NdeI
pCW"/2A6 hNPR
Bicistronic expression construct

Figure 1a(ii)

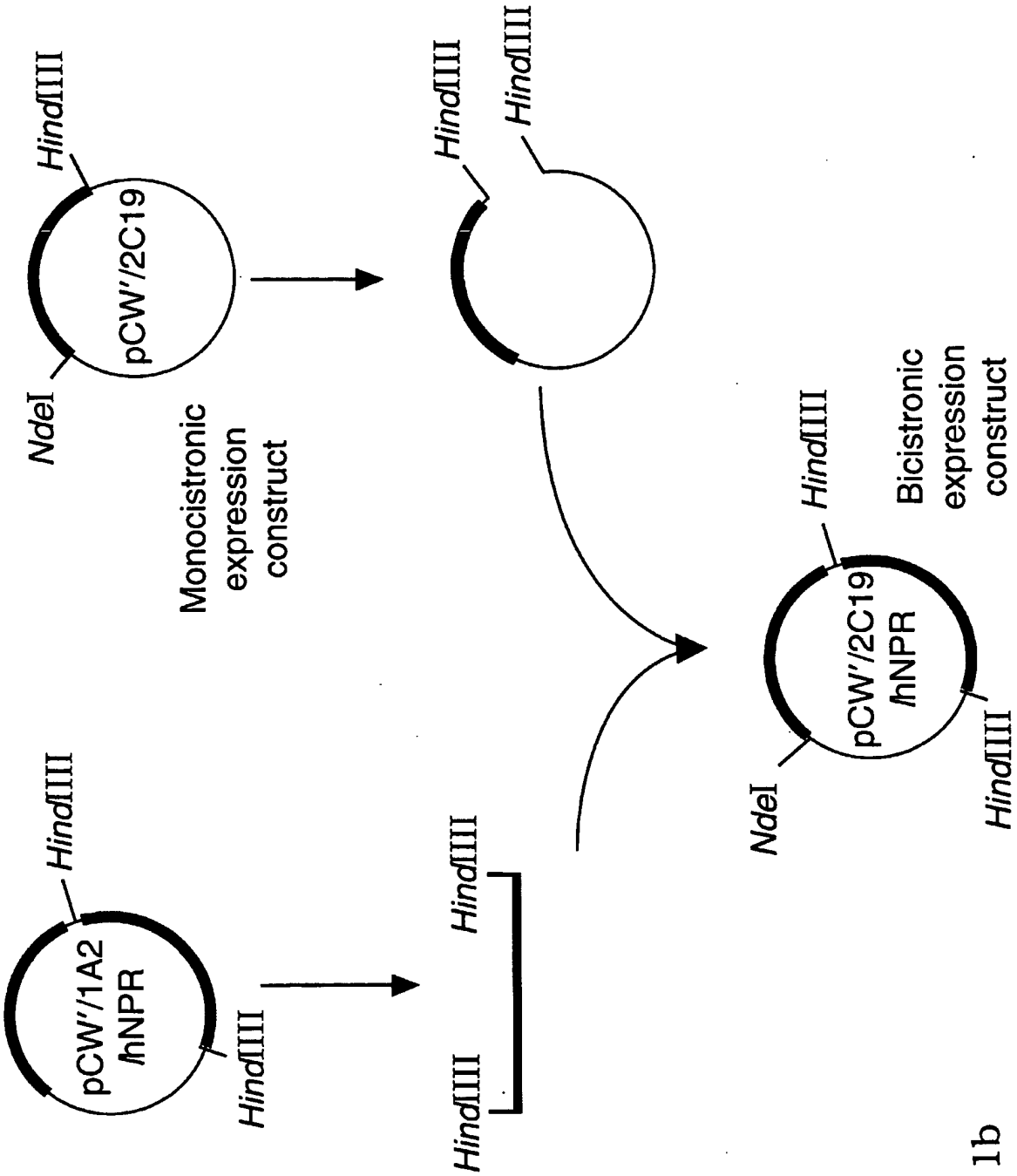
HindIII
pCW'/2C19
NdeI
Monocistronic expression construct
HindIII
HindIII
HindIII
pCW'/1A2 hNPR
HindIII
HindIII
HindIII
pCW'/2C19 hNPR
NdeI
HindIII
Bicistronic expression construct

Figure 1b

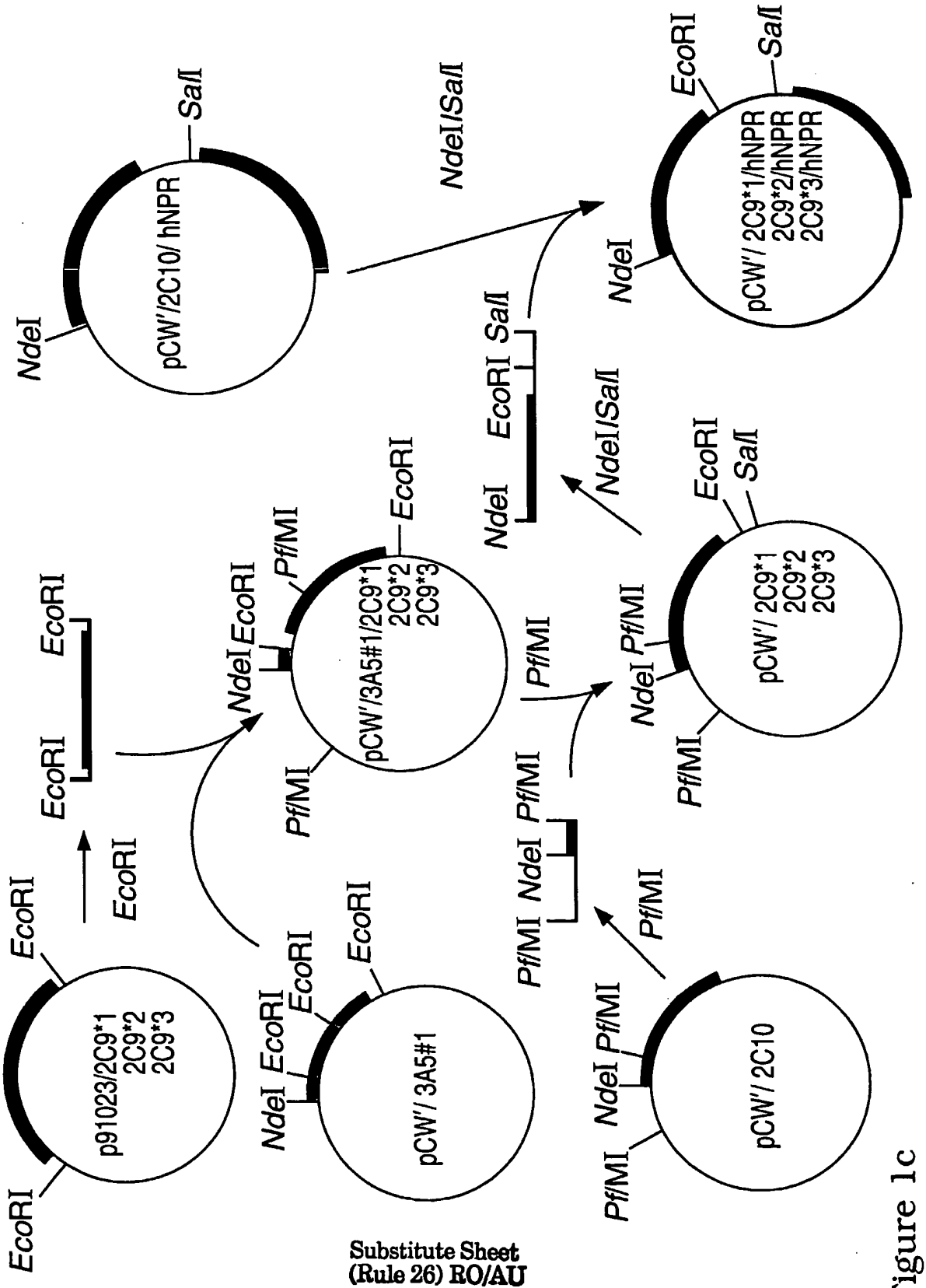
EcoRI
p91023/2C9*1
2C9*2
2C9*3
EcoRI
EcoRI
EcoRI
EcoRI
pCW'/ 3A5#1
NdeI
EcoRI
EcoRI
EcoRI
Pf/MI
pCW'/3A5#1/2C9*1
2C9*2
2C9*3
Pf/MI
NdeI
EcoRI
Pf/MI
Pf/MI
NdeI
Pf/MI
pCW'/ 2C10
NdeI
Pf/MI
Pf/MI
NdeI/SalI
NdeI
EcoRI
SalI
NdeI/SalI
NdeI
pCW'/2C10/ hNPR
SalI
NdeI
pCW'/ 2C9*1
2C9*2
2C9*3
EcoRI
SalI
NdeI
pCW'/ 2C9*1/hNPR
2C9*2/hNPR
2C9*3/hNPR
EcoRI
SalI

Figure 1c

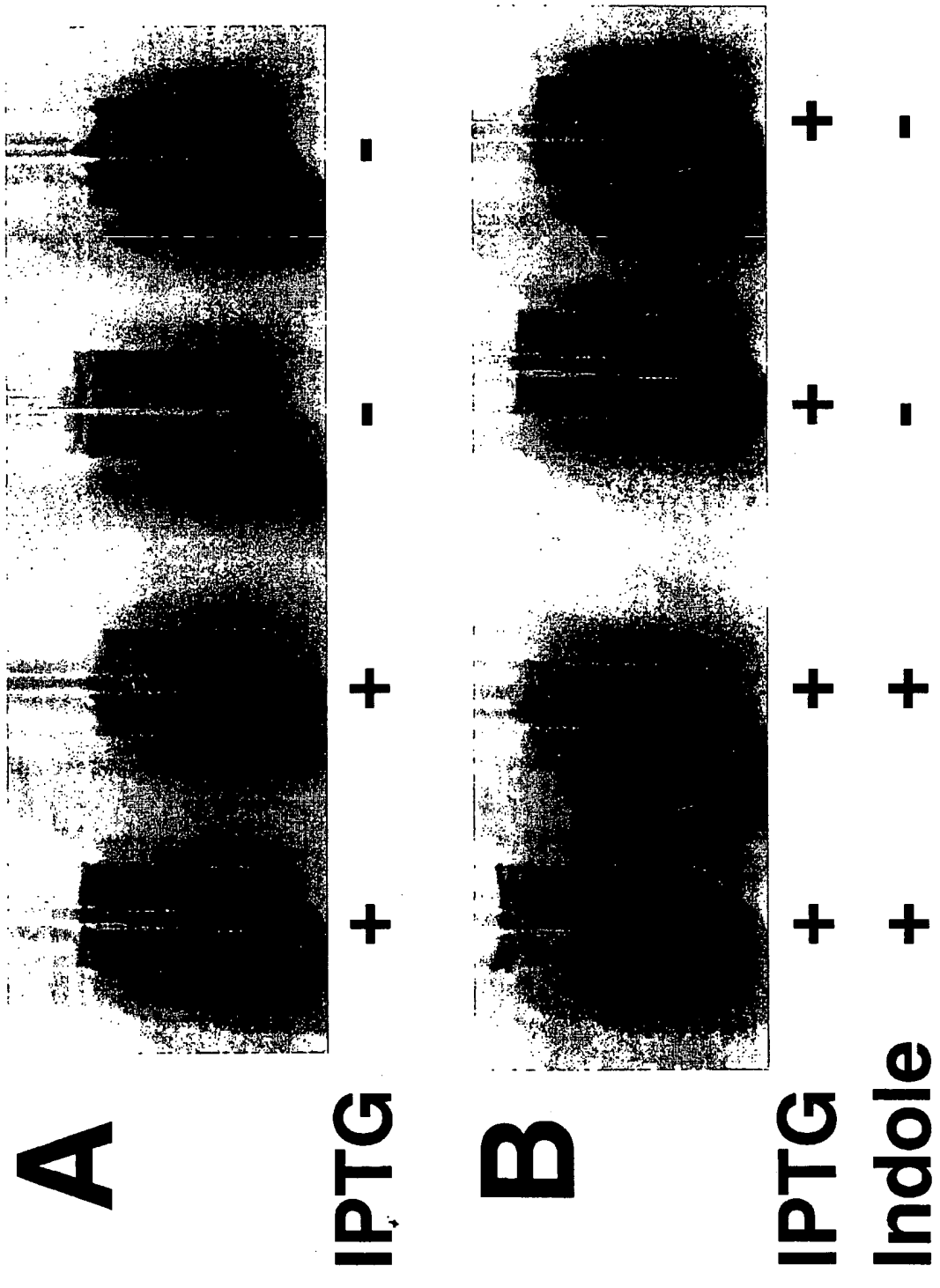

Figure 2

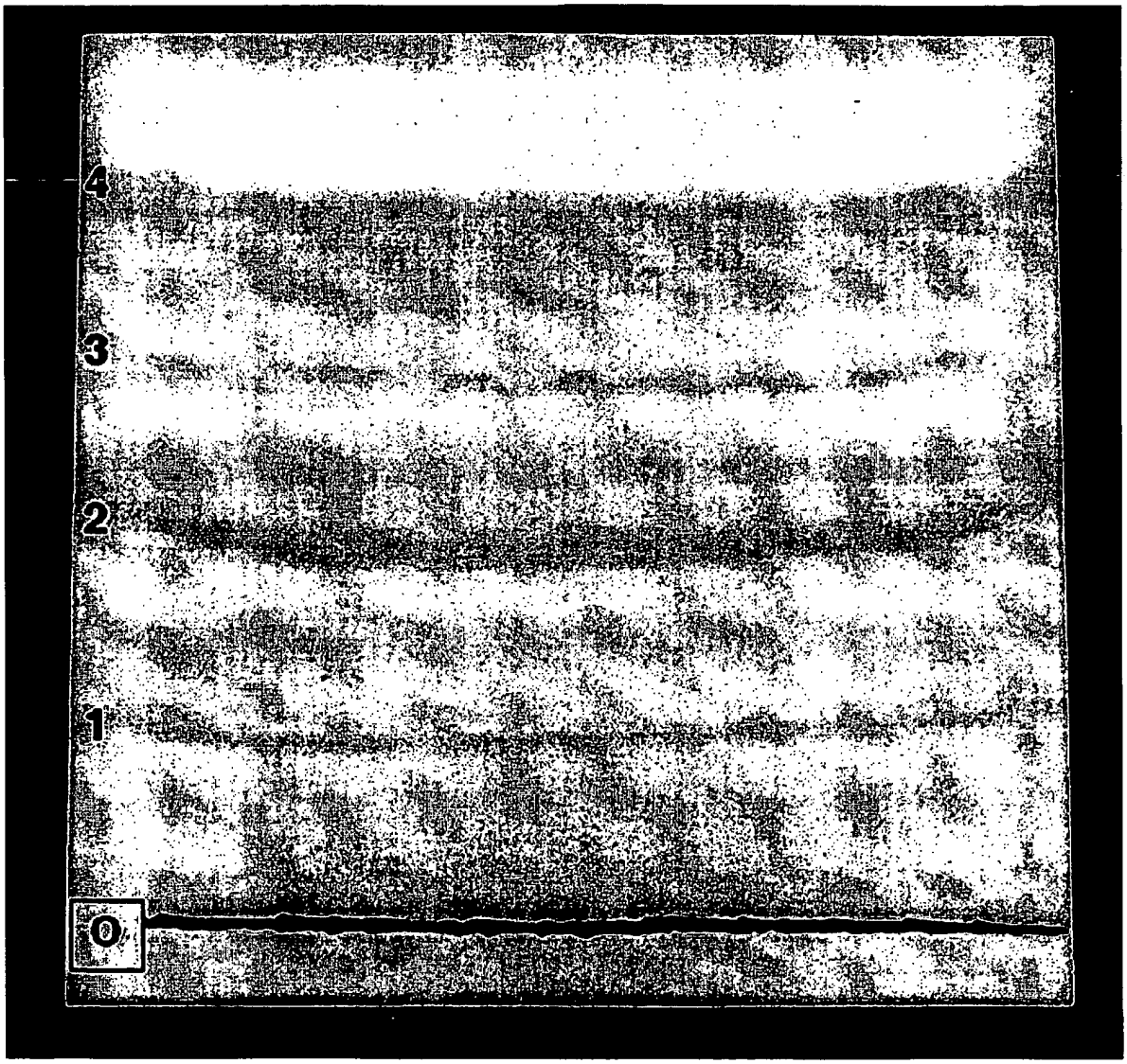

Figure 3

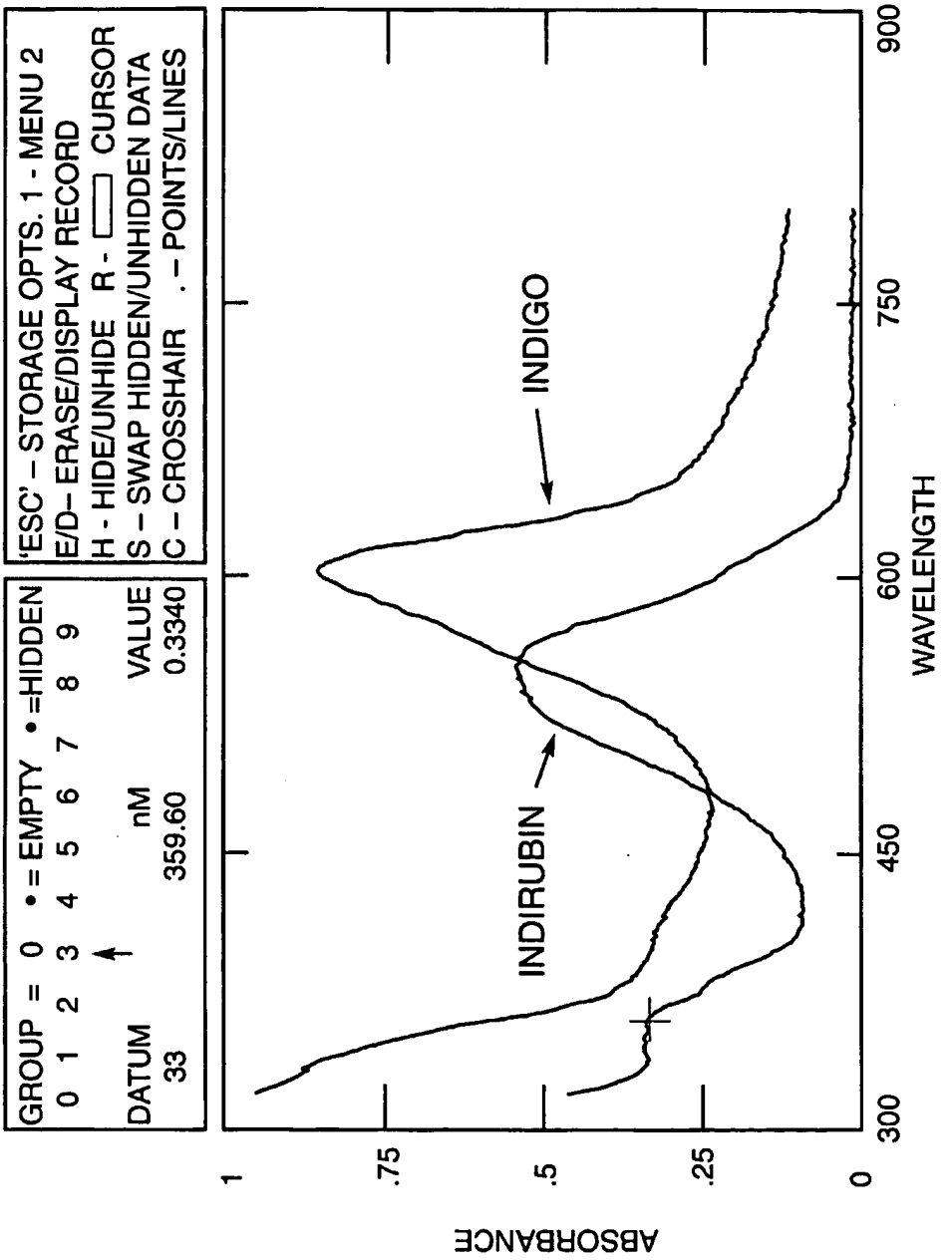
GROUP = 0 • = EMPTY • =HIDDEN
0 1 2 3 4 5 6 7 8 9
DATUM nM VALUE
33 359.60 0.3340
'ESC' – STORAGE OPTS. 1 - MENU 2
E/D– ERASE/DISPLAY RECORD
H - HIDE/UNHIDE R - CURSOR
S – SWAP HIDDEN/UNHIDDEN DATA
C – CROSSHAIR . – POINTS/LINES
INDIGO
INDIRUBIN
ABSORBANCE
0
.25
.5
.75
1
300
450
600
750
900
WAVELENGTH

Figure 4

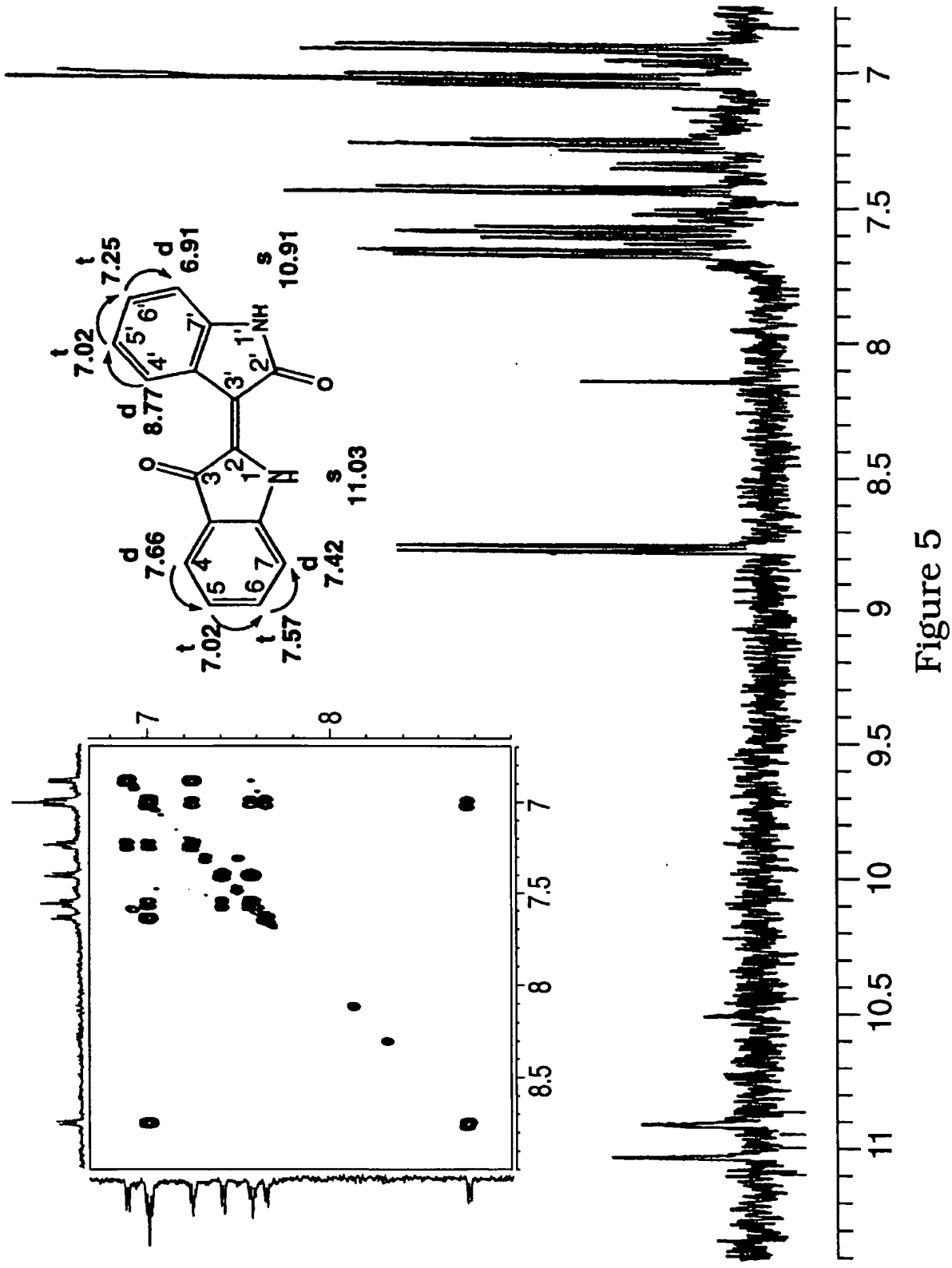
t 7.25
d 6.91
s 10.91
t 7.02
d 8.77
s 11.03
d 7.66
d 7.42
t 7.02
t 7.57
7
7.5
8
8.5
9
9.5
10
10.5
11

Figure 5

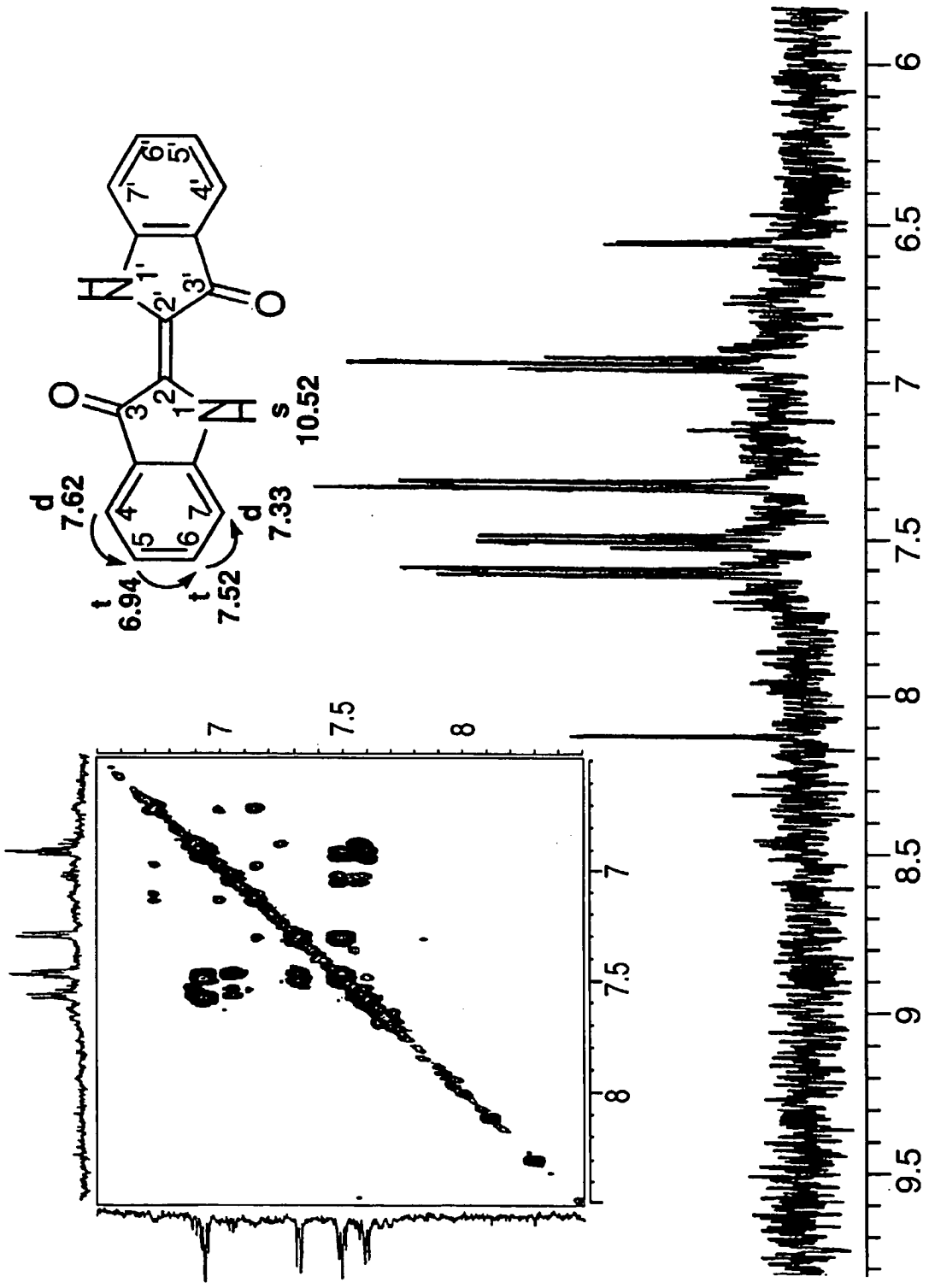

Figure 6

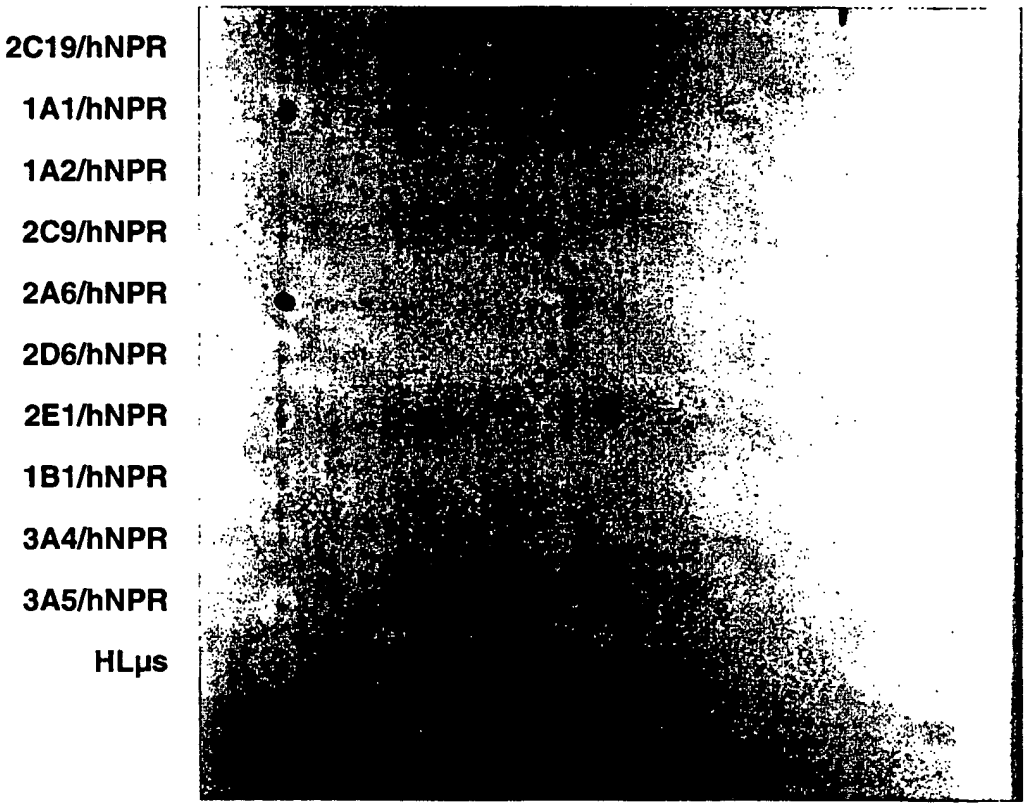
2C19/hNPR
1A1/hNPR
1A2/hNPR
2C9/hNPR
2A6/hNPR
2D6/hNPR
2E1/hNPR
1B1/hNPR
3A4/hNPR
3A5/hNPR
HLμs

Figure 7

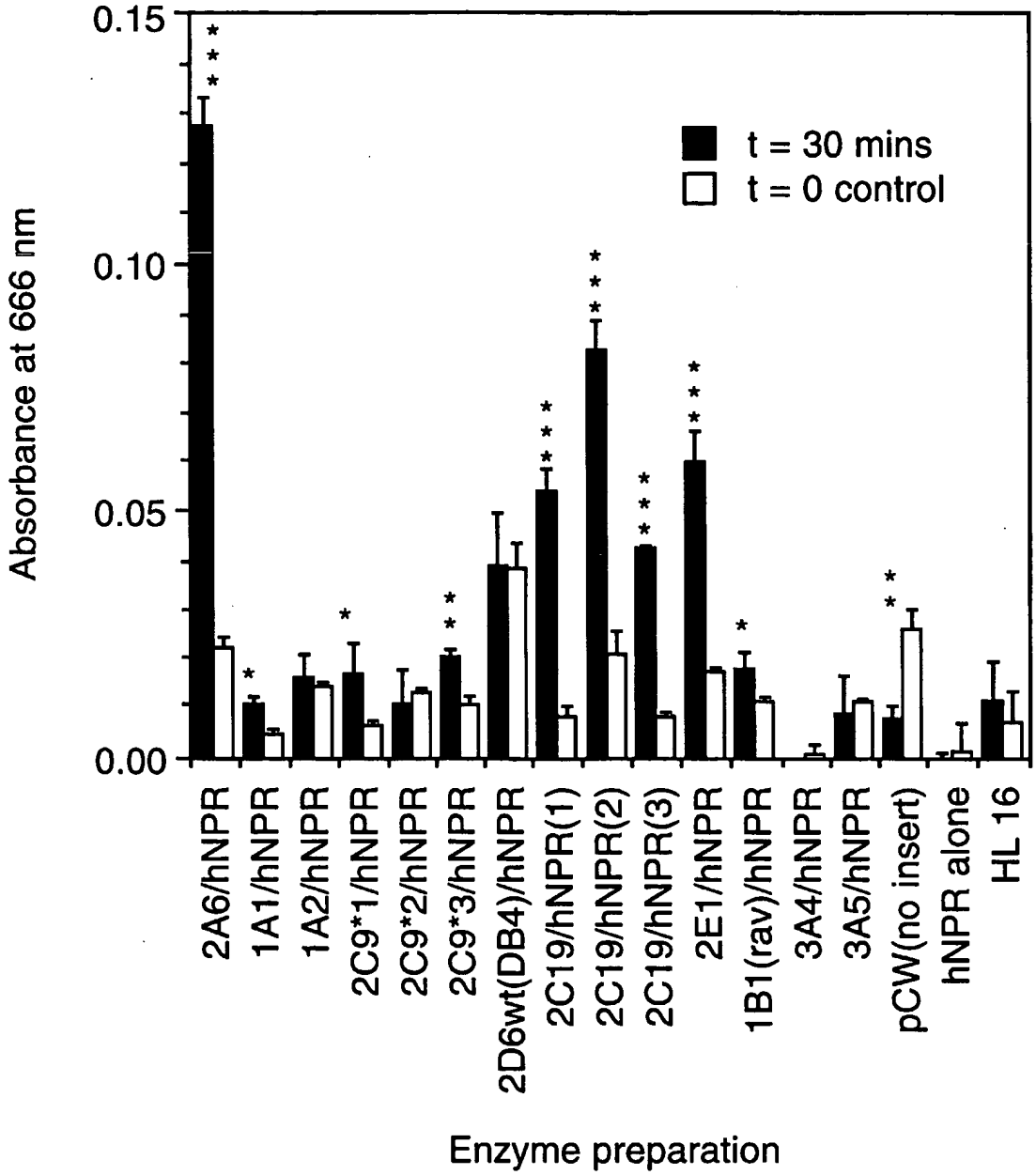
0.15
0.10
0.05
0.00
Absorbance at 666 nm
t = 30 mins
t = 0 control
2A6/hNPR
1A1/hNPR
1A2/hNPR
2C9*1/hNPR
2C9*2/hNPR
2C9*3/hNPR
2D6wt(DB4)/hNPR
2C19/hNPR(1)
2C19/hNPR(2)
2C19/hNPR(3)
2E1/hNPR
1B1(rav)/hNPR
3A4/hNPR
3A5/hNPR
pCW(no insert)
hNPR alone
HL 16
Enzyme preparation

Figure 8

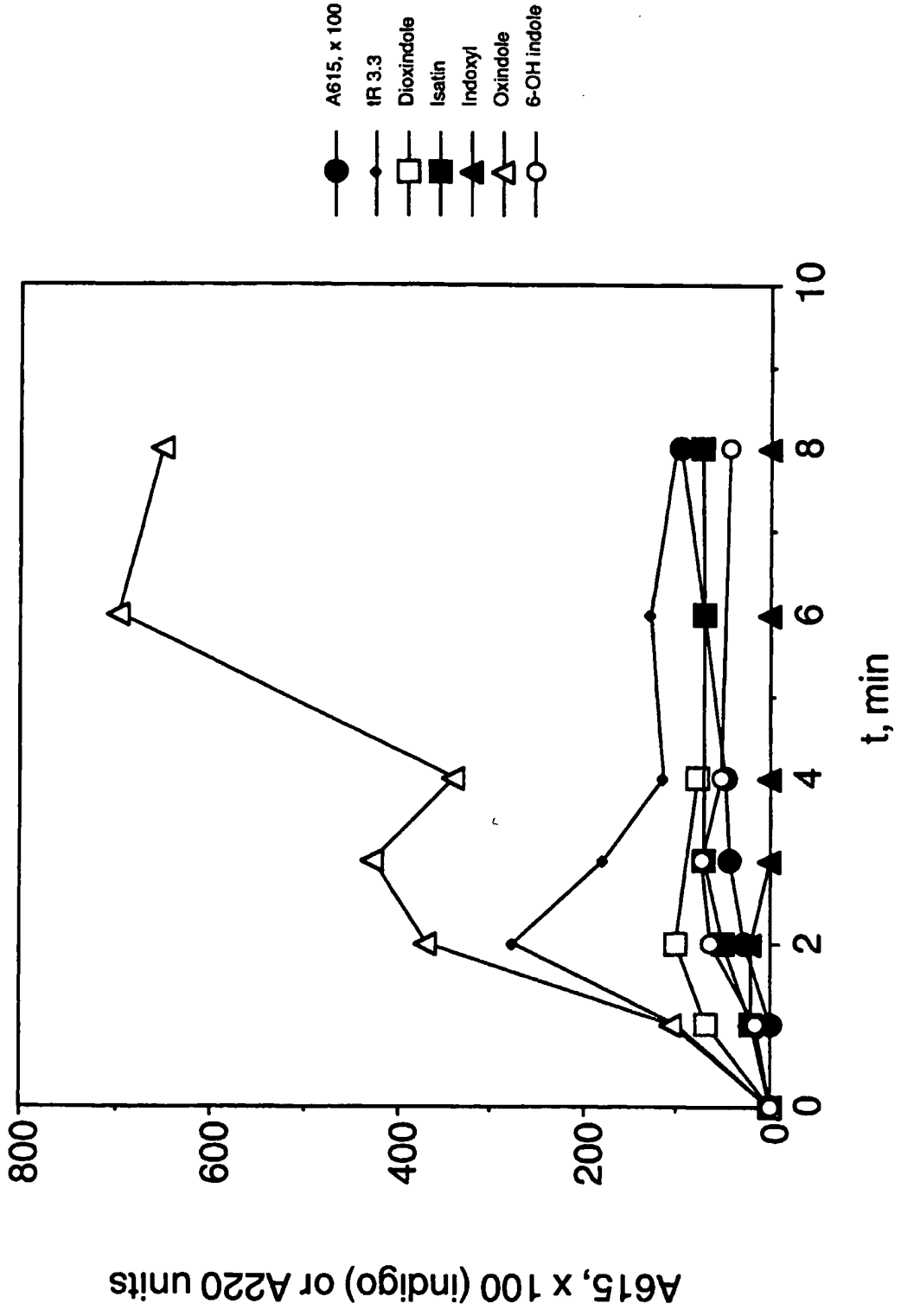
A615, x 100
tR 3.3
Dioxindole
Isatin
Indoxyl
Oxindole
6-OH indole
A615, x 100 (indigo) or A220 units
800
600
400
200
0
0
2
4
6
8
10
t, min

Figure 9

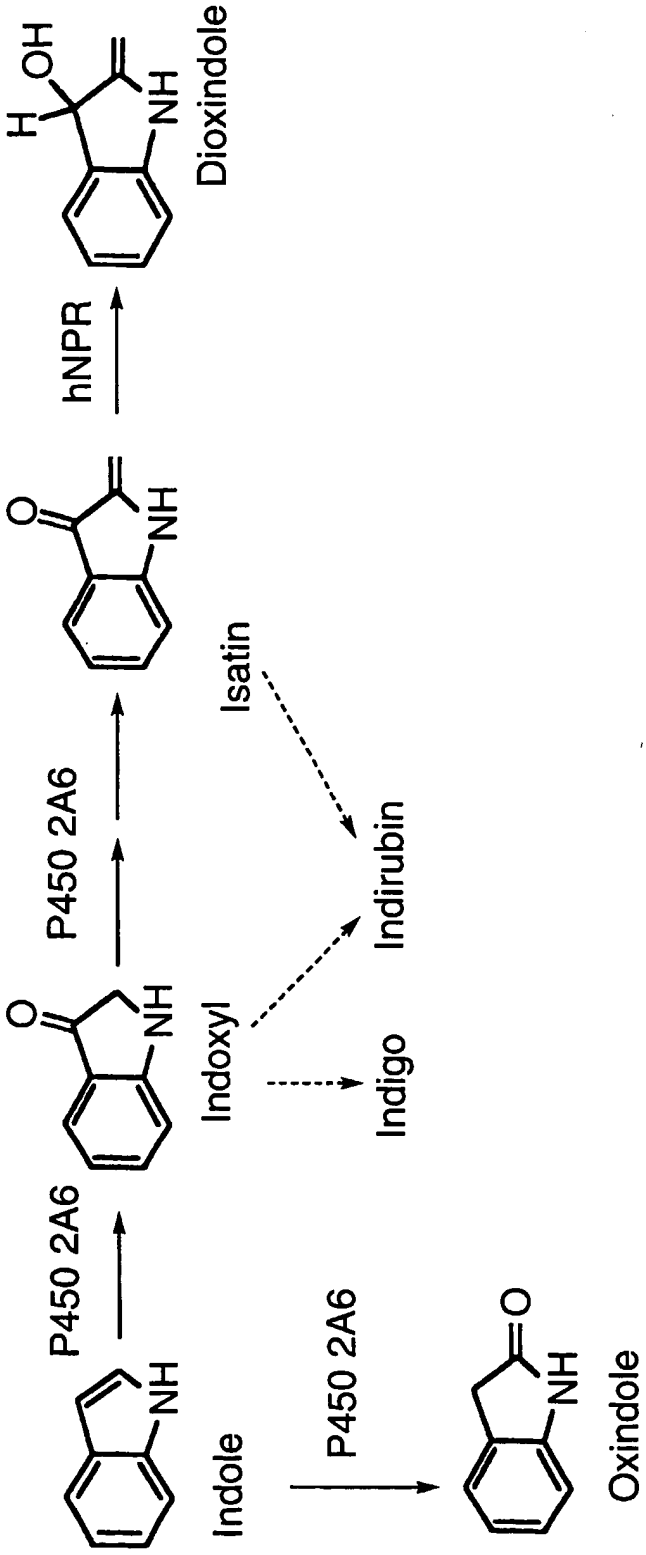
Indole
P450 2A6
Oxindole
P450 2A6
Indoxyl
Indigo
Isatin
Indirubin
P450 2A6
hNPR
Dioxindole
NH
O
OH
H

Figure 10

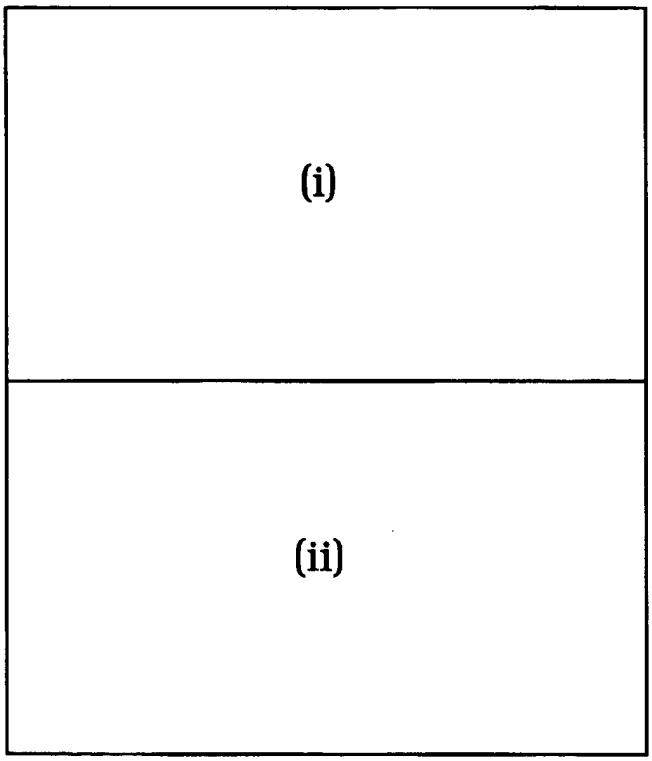

Figure 11

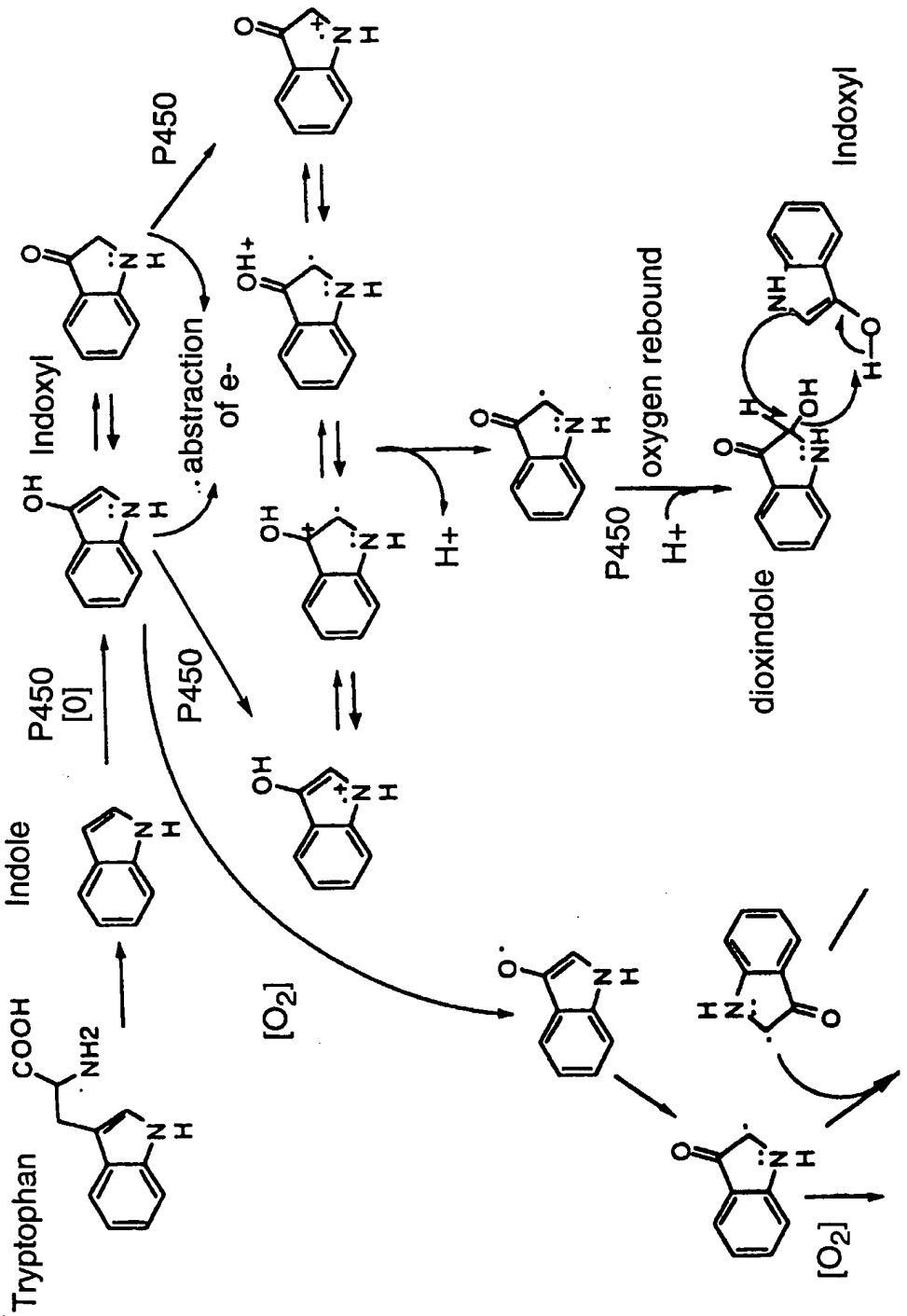
Tryptophan
COOH
NH2
Indole
P450
[O]
Indoxyl
P450
abstraction
of e-
OH+
H+
[O2]
oxygen rebound
P450
H+
Indoxyl
dioxindole
[O2]

Figure 11(i)

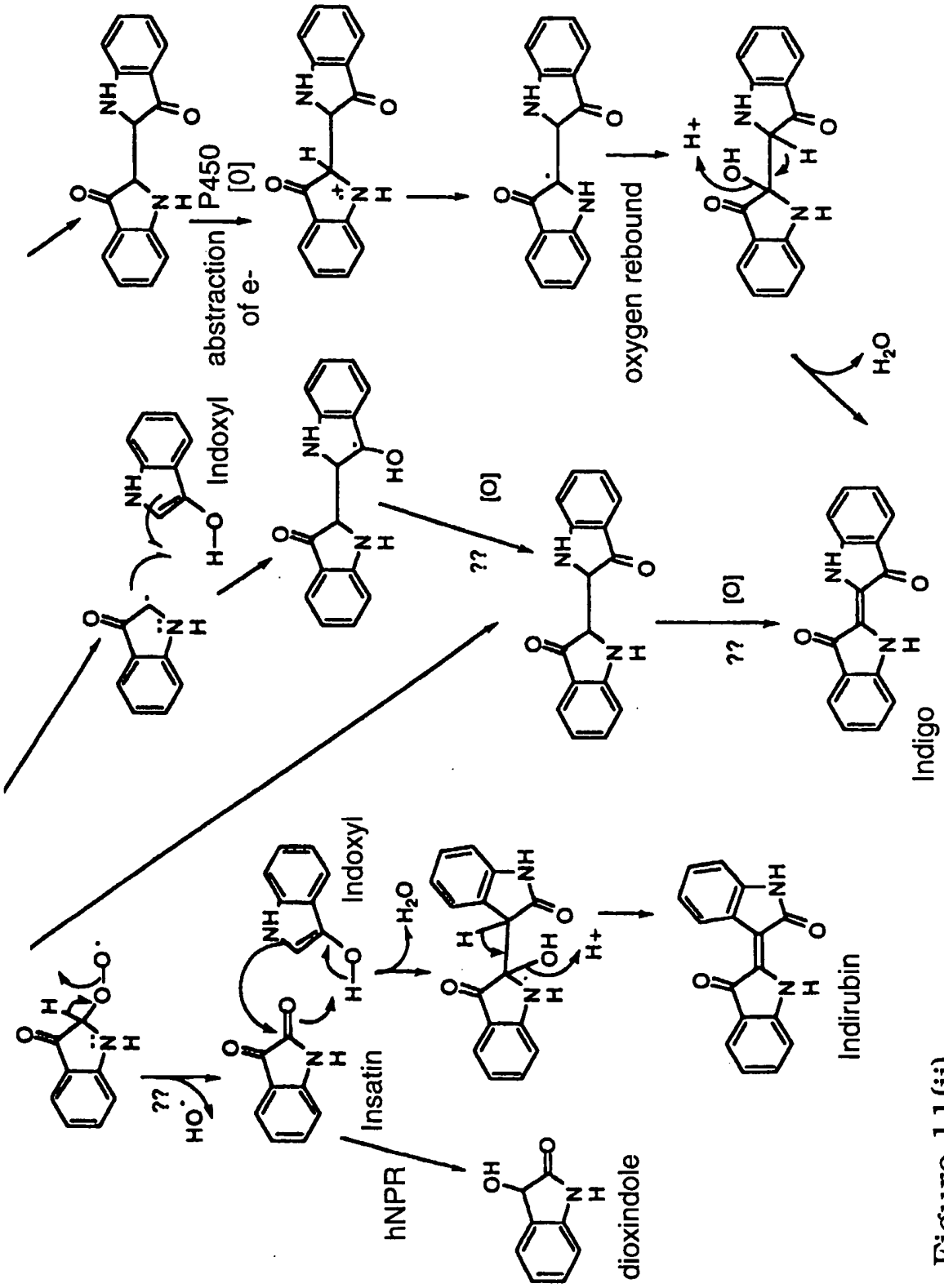
P450
[O]
abstraction of e-
oxygen rebound
H+
H2O
Indoxyl
[O]
??
[O]
??
Indigo
Indoxyl
H2O
??
HO
Insatin
hNPR
dioxindole
Indirubin

Figure 11(ii)

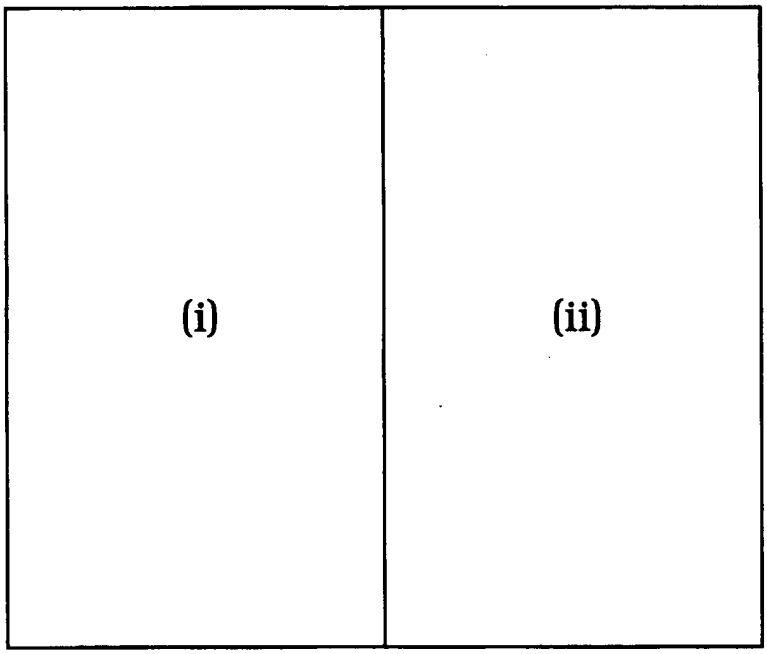

Figure 12

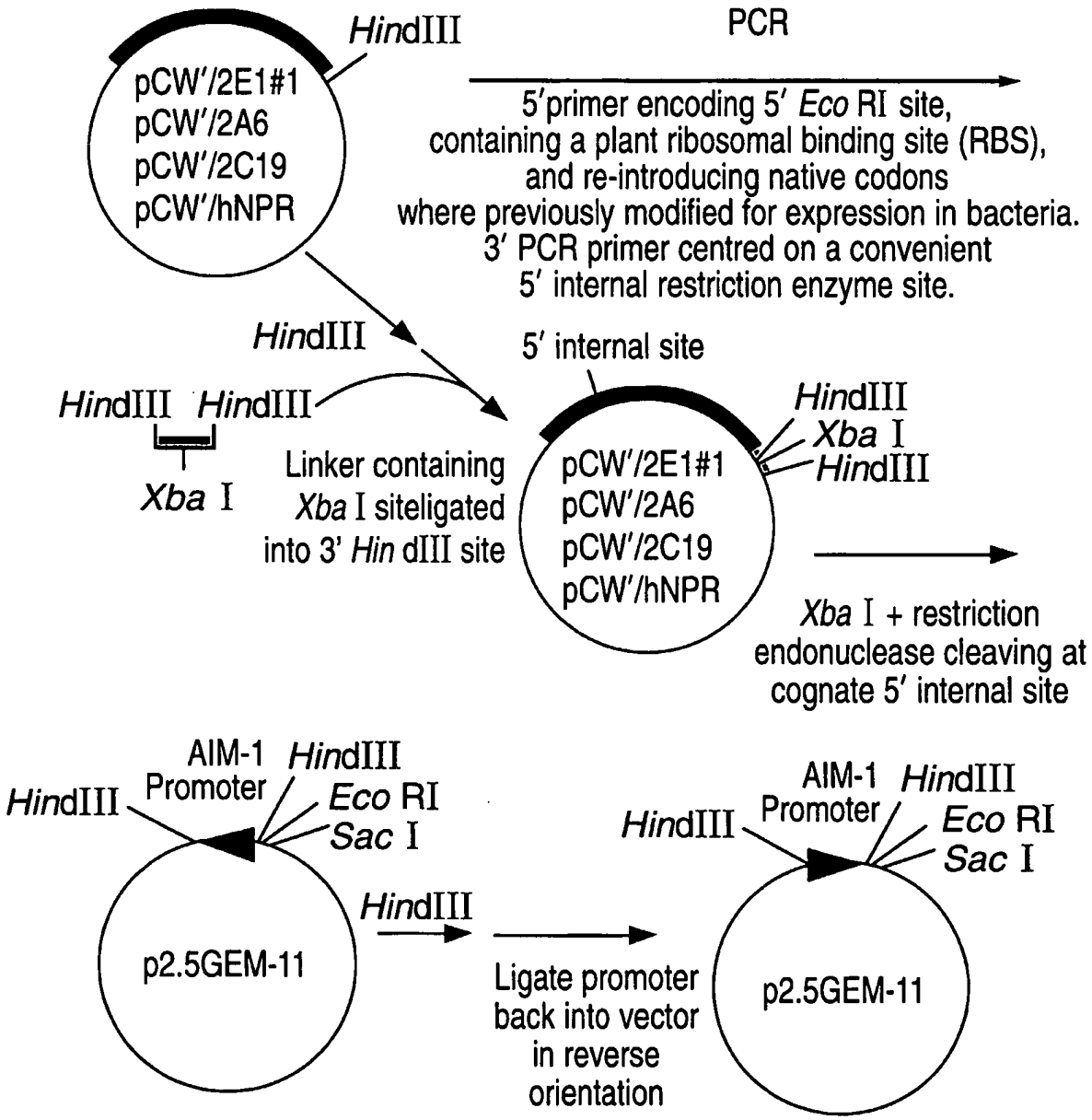
HindIII
PCR
pCW'/2E1#1
pCW'/2A6
pCW'/2C19
pCW'/hNPR
5'primer encoding 5' Eco RI site,
containing a plant ribosomal binding site (RBS),
and re-introducing native codons
where previously modified for expression in bacteria.
3' PCR primer centred on a convenient
5' internal restriction enzyme site.
HindIII
5' internal site
HindIII HindIII
Xba I
Linker containing
Xba I siteligated
into 3' Hin dIII site
HindIII
Xba I
HindIII
pCW'/2E1#1
pCW'/2A6
pCW'/2C19
pCW'/hNPR
Xba I + restriction
endonuclease cleaving at
cognate 5' internal site
AIM-1
Promoter
HindIII
HindIII
Eco RI
Sac I
HindIII
p2.5GEM-11
Ligate promoter
back into vector
in reverse
orientation
AIM-1
Promoter
HindIII
HindIII
Eco RI
Sac I
p2.5GEM-11

Figure 12(i)

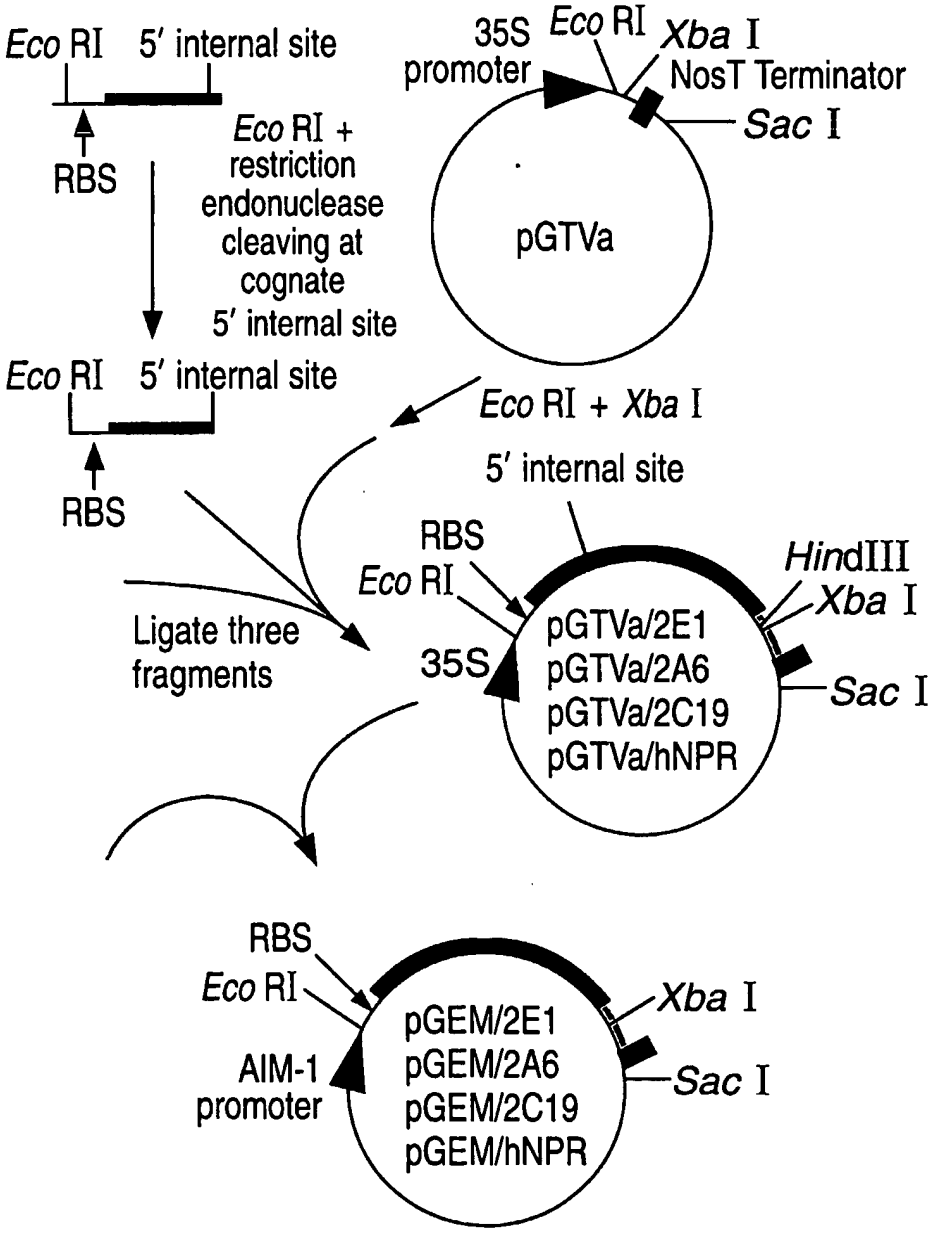
Eco RI
5′ internal site
RBS
Eco RI +
restriction
endonuclease
cleaving at
cognate
5′ internal site
Eco RI
5′ internal site
RBS
35S
promoter
Eco RI
Xba I
NosT Terminator
Sac I
pGTVa
Eco RI + Xba I
5′ internal site
RBS
Eco RI
35S
Ligate three
fragments
pGTVa/2E1
pGTVa/2A6
pGTVa/2C19
pGTVa/hNPR
HindIII
Xba I
Sac I
RBS
Eco RI
AIM-1
promoter
pGEM/2E1
pGEM/2A6
pGEM/2C19
pGEM/hNPR
Xba I
Sac I

Figure 12(ii)

SEQUENCE LISTING

<110> The University of Queensland, Vanderbilt University

<120> Pigment production

<130> 2302839/EJH

<140> International
<141> 2000-08-11

<150> PQ2156
<151> 1999-08-11

<160> 22

<170> PatentIn Ver. 2.1

<210> 1
<211> 1485
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1482)

<400> 1

```
atg gct gct tca gga atg tta tta gtg gcc ttg ctg gtc tgc ctg act   48
Met Ala Ala Ser Gly Met Leu Leu Val Ala Leu Leu Val Cys Leu Thr
  1               5                  10                  15

gtg atg gtc ttg atg tct gtt tgg cag cag agg aag agc aag ggg aag   96
Val Met Val Leu Met Ser Val Trp Gln Gln Arg Lys Ser Lys Gly Lys
             20                  25                  30

ctg cct ccg gga ccc acc cca ttg ccc ttc att gga aac tac ctg cag   144
Leu Pro Pro Gly Pro Thr Pro Leu Pro Phe Ile Gly Asn Tyr Leu Gln
         35                  40                  45
```

```
ctg aac aca gag cag atg tac aac tcc ctc atg aag atc agt gag cgc   192
Leu Asn Thr Glu Gln Met Tyr Asn Ser Leu Met Lys Ile Ser Glu Arg
     50                  55                  60

tat ggc ccc gtg ttc acc att cac ttg ggg ccc cgg cgg gtc gtg gtg   240
Tyr Gly Pro Val Phe Thr Ile His Leu Gly Pro Arg Arg Val Val Val
 65                  70                  75                  80

ctg tgt gga cat gat gcc gtc agg gag gct ctg gtg gac cag gct gag   288
Leu Cys Gly His Asp Ala Val Arg Glu Ala Leu Val Asp Gln Ala Glu
                 85                  90                  95

gag ttc agc ggg cga ggc gag caa gcc acc ttc gac tgg gtc ttc aaa   336
Glu Phe Ser Gly Arg Gly Glu Gln Ala Thr Phe Asp Trp Val Phe Lys
            100                 105                 110

ggc tat ggc gtg gta ttc agc aac ggg gag cgc gcc aag cag ctc cgg   384
Gly Tyr Gly Val Val Phe Ser Asn Gly Glu Arg Ala Lys Gln Leu Arg
        115                 120                 125

cgc ttc tcc atc gcc acc ctg cgg gac ttc ggg gtg ggc aag cga ggc   432
Arg Phe Ser Ile Ala Thr Leu Arg Asp Phe Gly Val Gly Lys Arg Gly
    130                 135                 140

atc gag gag cgc atc cag gag gag gcg ggc ttc ctc atc gac gcc ctc   480
Ile Glu Glu Arg Ile Gln Glu Glu Ala Gly Phe Leu Ile Asp Ala Leu
145                 150                 155                 160

cgg ggc act ggc ggc gcc aat atc gat ccc acc ttc ttc ctg agc cgc   528
Arg Gly Thr Gly Gly Ala Asn Ile Asp Pro Thr Phe Phe Leu Ser Arg
                165                 170                 175

aca gtc tcc aat gtc atc agc tcc att gtc ttt ggg gac cgc ttt gac   576
Thr Val Ser Asn Val Ile Ser Ser Ile Val Phe Gly Asp Arg Phe Asp
            180                 185                 190

tat aag gac aaa gag ttc ctg tca ctg ttg cgc atg atg cta gga atc   624
Tyr Lys Asp Lys Glu Phe Leu Ser Leu Leu Arg Met Met Leu Gly Ile
```

```
        195                 200                 205

ttc cag ttc acg tca acc tcc acg ggg cag ctc tat gag atg ttc tct   672
Phe Gln Phe Thr Ser Thr Ser Thr Gly Gln Leu Tyr Glu Met Phe Ser
    210                 215                 220

tcg gtg atg aaa cac ctg cca gga cca cag caa cag gcc ttt cag ttg   720
Ser Val Met Lys His Leu Pro Gly Pro Gln Gln Gln Ala Phe Gln Leu
225                 230                 235                 240

ctg caa ggg ctg gag gac ttc ata gcc aag aag gtg gag cac aac cag   768
Leu Gln Gly Leu Glu Asp Phe Ile Ala Lys Lys Val Glu His Asn Gln
                245                 250                 255

cgc acg ctg gat ccc aat tcc cca cgg gac ttc att gac tcc ttt ctc   816
Arg Thr Leu Asp Pro Asn Ser Pro Arg Asp Phe Ile Asp Ser Phe Leu
            260                 265                 270

atc cgc atg cag gag gag gag aag aac ccc aac acg gag ttc tac ttg   864
Ile Arg Met Gln Glu Glu Glu Lys Asn Pro Asn Thr Glu Phe Tyr Leu
        275                 280                 285

aaa aac ctg gtg atg acc acg ttg aac ctc ttc att ggg ggc acc gag   912
Lys Asn Leu Val Met Thr Thr Leu Asn Leu Phe Ile Gly Gly Thr Glu
    290                 295                 300

acc gtc agc acc acc ctg cgc tat ggc ttc ttg ctg ctc atg aag cac   960
Thr Val Ser Thr Thr Leu Arg Tyr Gly Phe Leu Leu Leu Met Lys His
305                 310                 315                 320

cca gag gtg gag gcc aag gtc cat gag gag att gac aga gtg atc ggc   1008
Pro Glu Val Glu Ala Lys Val His Glu Glu Ile Asp Arg Val Ile Gly
                325                 330                 335

aag aac cgg cag ccc aag ttt gag gac cgg gcc aag atg ccc tac atg   1056
Lys Asn Arg Gln Pro Lys Phe Glu Asp Arg Ala Lys Met Pro Tyr Met
            340                 345                 350

gag gca gtg atc cac gag atc caa aga ttt gga gac gtg atc ccc atg   1104
```

```
Glu Ala Val Ile His Glu Ile Gln Arg Phe Gly Asp Val Ile Pro Met
        355                 360                 365

agt ttg gcc cgc aga gtc aaa aag gac acc aag ttt cgg gat ttc ttc   1152
Ser Leu Ala Arg Arg Val Lys Lys Asp Thr Lys Phe Arg Asp Phe Phe
    370                 375                 380

ctc cct aag ggc acc gaa gtg tac cct atg ctg ggc tct gtg ctg aga   1200
Leu Pro Lys Gly Thr Glu Val Tyr Pro Met Leu Gly Ser Val Leu Arg
385                 390                 395                 400

gac ccc agt ttc ttc tcc aac ccc cag gac ttc aat ccc cag cac ttc   1248
Asp Pro Ser Phe Phe Ser Asn Pro Gln Asp Phe Asn Pro Gln His Phe
                405                 410                 415

ctg aat gag aag ggg cag ttt aag aag agt gat gct ttt gtg ccc ttt   1296
Leu Asn Glu Lys Gly Gln Phe Lys Lys Ser Asp Ala Phe Val Pro Phe
            420                 425                 430

tcc atc gga aag cgg aac tgt ttc gga gaa ggc ctg gcc aga atg gag   1344
Ser Ile Gly Lys Arg Asn Cys Phe Gly Glu Gly Leu Ala Arg Met Glu
        435                 440                 445

ctc ttt ctc ttc ttc acc acc gtc atg cag aac ttc cgc ctc aag tcc   1392
Leu Phe Leu Phe Phe Thr Thr Val Met Gln Asn Phe Arg Leu Lys Ser
    450                 455                 460

tcc cag tca cct aag gac att gac gtg tcc ccc aaa cac gtg ggc ttt   1440
Ser Gln Ser Pro Lys Asp Ile Asp Val Ser Pro Lys His Val Gly Phe
465                 470                 475                 480

gcc acg atc cca cga aac tac acc atg agc ttc ctg ccc cgc tga       1485
Ala Thr Ile Pro Arg Asn Tyr Thr Met Ser Phe Leu Pro Arg
                485                 490
```

<210> 2
<211> 494
<212> PRT

<213> Human cells

<400> 2

```
Met Ala Ala Ser Gly Met Leu Leu Val Ala Leu Leu Val Cys Leu Thr
  1               5                  10                  15

Val Met Val Leu Met Ser Val Trp Gln Gln Arg Lys Ser Lys Gly Lys
             20                  25                  30

Leu Pro Pro Gly Pro Thr Pro Leu Pro Phe Ile Gly Asn Tyr Leu Gln
         35                  40                  45

Leu Asn Thr Glu Gln Met Tyr Asn Ser Leu Met Lys Ile Ser Glu Arg
     50                  55                  60

Tyr Gly Pro Val Phe Thr Ile His Leu Gly Pro Arg Arg Val Val Val
 65                  70                  75                  80

Leu Cys Gly His Asp Ala Val Arg Glu Ala Leu Val Asp Gln Ala Glu
                 85                  90                  95

Glu Phe Ser Gly Arg Gly Glu Gln Ala Thr Phe Asp Trp Val Phe Lys
            100                 105                 110

Gly Tyr Gly Val Val Phe Ser Asn Gly Glu Arg Ala Lys Gln Leu Arg
        115                 120                 125

Arg Phe Ser Ile Ala Thr Leu Arg Asp Phe Gly Val Gly Lys Arg Gly
    130                 135                 140

Ile Glu Glu Arg Ile Gln Glu Glu Ala Gly Phe Leu Ile Asp Ala Leu
145                 150                 155                 160

Arg Gly Thr Gly Gly Ala Asn Ile Asp Pro Thr Phe Phe Leu Ser Arg
                165                 170                 175

Thr Val Ser Asn Val Ile Ser Ser Ile Val Phe Gly Asp Arg Phe Asp
            180                 185                 190
```

```
Tyr Lys Asp Lys Glu Phe Leu Ser Leu Leu Arg Met Met Leu Gly Ile
        195                 200                 205

Phe Gln Phe Thr Ser Thr Ser Thr Gly Gln Leu Tyr Glu Met Phe Ser
    210                 215                 220

Ser Val Met Lys His Leu Pro Gly Pro Gln Gln Gln Ala Phe Gln Leu
225                 230                 235                 240

Leu Gln Gly Leu Glu Asp Phe Ile Ala Lys Lys Val Glu His Asn Gln
                245                 250                 255

Arg Thr Leu Asp Pro Asn Ser Pro Arg Asp Phe Ile Asp Ser Phe Leu
            260                 265                 270

Ile Arg Met Gln Glu Glu Glu Lys Asn Pro Asn Thr Glu Phe Tyr Leu
        275                 280                 285

Lys Asn Leu Val Met Thr Thr Leu Asn Leu Phe Ile Gly Gly Thr Glu
    290                 295                 300

Thr Val Ser Thr Thr Leu Arg Tyr Gly Phe Leu Leu Leu Met Lys His
305                 310                 315                 320

Pro Glu Val Glu Ala Lys Val His Glu Glu Ile Asp Arg Val Ile Gly
                325                 330                 335

Lys Asn Arg Gln Pro Lys Phe Glu Asp Arg Ala Lys Met Pro Tyr Met
            340                 345                 350

Glu Ala Val Ile His Glu Ile Gln Arg Phe Gly Asp Val Ile Pro Met
        355                 360                 365

Ser Leu Ala Arg Arg Val Lys Lys Asp Thr Lys Phe Arg Asp Phe Phe
    370                 375                 380

Leu Pro Lys Gly Thr Glu Val Tyr Pro Met Leu Gly Ser Val Leu Arg
385                 390                 395                 400
```

```
Asp Pro Ser Phe Phe Ser Asn Pro Gln Asp Phe Asn Pro Gln His Phe
                405                 410                 415

Leu Asn Glu Lys Gly Gln Phe Lys Lys Ser Asp Ala Phe Val Pro Phe
            420                 425                 430

Ser Ile Gly Lys Arg Asn Cys Phe Gly Glu Gly Leu Ala Arg Met Glu
        435                 440                 445

Leu Phe Leu Phe Phe Thr Thr Val Met Gln Asn Phe Arg Leu Lys Ser
    450                 455                 460

Ser Gln Ser Pro Lys Asp Ile Asp Val Ser Pro Lys His Val Gly Phe
465                 470                 475                 480

Ala Thr Ile Pro Arg Asn Tyr Thr Met Ser Phe Leu Pro Arg
                485                 490
```

<210> 3
<211> 1419
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1416)

<400> 3

```
atg gct cgt caa gtt cat tct tct tgg aat ctg ccc cca ggc cct ttc   48
Met Ala Arg Gln Val His Ser Ser Trp Asn Leu Pro Pro Gly Pro Phe
  1               5                  10                  15

ccg ctt ccc atc atc ggg aac ctc ttc cag ttg gaa ttg aag aat att   96
Pro Leu Pro Ile Ile Gly Asn Leu Phe Gln Leu Glu Leu Lys Asn Ile
            20                  25                  30

ccc aag tcc ttc acc cgg ttg gcc cag cgc ttc ggg ccg gtg ttc acg   144
```

```
Pro Lys Ser Phe Thr Arg Leu Ala Gln Arg Phe Gly Pro Val Phe Thr
         35                  40                  45

ctg tac gtg ggc tcg cag cgc atg gtg gtg atg cac ggc tac aag gcg  192
Leu Tyr Val Gly Ser Gln Arg Met Val Val Met His Gly Tyr Lys Ala
     50                  55                  60

gtg aag gaa gcg ctg ctg gac tac aag gac gag ttc tcg ggc aga ggc  240
Val Lys Glu Ala Leu Leu Asp Tyr Lys Asp Glu Phe Ser Gly Arg Gly
 65                  70                  75                  80

gac ctc ccc gcg ttc cat gcg cac agg gac agg gga atc att ttt aat  288
Asp Leu Pro Ala Phe His Ala His Arg Asp Arg Gly Ile Ile Phe Asn
                 85                  90                  95

aat gga cct acc tgg aag gac atc cgg cgg ttt tcc ctg acc acc ctc  336
Asn Gly Pro Thr Trp Lys Asp Ile Arg Arg Phe Ser Leu Thr Thr Leu
            100                 105                 110

cgg aac tat ggg atg ggg aaa cag ggc aat gag agc cgg atc cag agg  384
Arg Asn Tyr Gly Met Gly Lys Gln Gly Asn Glu Ser Arg Ile Gln Arg
        115                 120                 125

gag gcc cac ttc ctg ctg gaa gca ctc agg aag acc caa ggc cag cct  432
Glu Ala His Phe Leu Leu Glu Ala Leu Arg Lys Thr Gln Gly Gln Pro
    130                 135                 140

ttc gac ccc acc ttc ctc atc ggc tgc gcg ccc tgc aac gtc ata gcc  480
Phe Asp Pro Thr Phe Leu Ile Gly Cys Ala Pro Cys Asn Val Ile Ala
145                 150                 155                 160

gac atc ctc ttc cgc aag cat ttt gac tac aat gat gag aag ttt cta  528
Asp Ile Leu Phe Arg Lys His Phe Asp Tyr Asn Asp Glu Lys Phe Leu
                165                 170                 175

agg ctg atg tat ttg ttt aat gag aac ttc cac cta ctc agc act ccc  576
Arg Leu Met Tyr Leu Phe Asn Glu Asn Phe His Leu Leu Ser Thr Pro
            180                 185                 190
```

```
tgg ctc cag ctt tac aat aat ttt ccc agc ttt cta cac tac ttg cct   624
Trp Leu Gln Leu Tyr Asn Asn Phe Pro Ser Phe Leu His Tyr Leu Pro
        195                 200                 205

gga agc cac aga aaa gtc ata aaa aat gtg gct gaa gta aaa gag tat   672
Gly Ser His Arg Lys Val Ile Lys Asn Val Ala Glu Val Lys Glu Tyr
    210                 215                 220

gtg tct gaa agg gtg aag gag cac cat caa tct ctg gac ccc aac tgt   720
Val Ser Glu Arg Val Lys Glu His His Gln Ser Leu Asp Pro Asn Cys
225                 230                 235                 240

ccc cgg gac ctc acc gac tgc ctg ctc gtg gaa atg gag aag gaa aag   768
Pro Arg Asp Leu Thr Asp Cys Leu Leu Val Glu Met Glu Lys Glu Lys
                245                 250                 255

cac agt gca gag cgc ttg tac aca atg gac ggt atc acc gtg act gtg   816
His Ser Ala Glu Arg Leu Tyr Thr Met Asp Gly Ile Thr Val Thr Val
            260                 265                 270

gcc gac ctg ttc ttt gcg ggg aca gag acc acc agc aca act ctg aga   864
Ala Asp Leu Phe Phe Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu Arg
        275                 280                 285

tat ggg ctc ctg att ctc atg aaa tac cct gag atc gaa gag aag ctc   912
Tyr Gly Leu Leu Ile Leu Met Lys Tyr Pro Glu Ile Glu Glu Lys Leu
    290                 295                 300

cat gaa gaa att gac agg gtg att ggg cca agc cga atc cct gcc atc   960
His Glu Glu Ile Asp Arg Val Ile Gly Pro Ser Arg Ile Pro Ala Ile
305                 310                 315                 320

aag gat agg caa gag atg ccc tac atg gat gct gtg gtg cat gag att  1008
Lys Asp Arg Gln Glu Met Pro Tyr Met Asp Ala Val Val His Glu Ile
                325                 330                 335

cag cgg ttc atc acc ctc gtg ccc tcc aac ctg ccc cat gaa gca acc  1056
Gln Arg Phe Ile Thr Leu Val Pro Ser Asn Leu Pro His Glu Ala Thr
            340                 345                 350
```

```
cga gac acc att ttc aga gga tac ctc atc ccc aag ggc aca gtc gta   1104
Arg Asp Thr Ile Phe Arg Gly Tyr Leu Ile Pro Lys Gly Thr Val Val
        355                 360                 365

gtg cca act ctg gac tct gtt ttg tat gac aac caa gaa ttt cct gat   1152
Val Pro Thr Leu Asp Ser Val Leu Tyr Asp Asn Gln Glu Phe Pro Asp
    370                 375                 380

cca gaa aag ttt aag cca gaa cac ttc ctg aat gaa aat gga aag ttc   1200
Pro Glu Lys Phe Lys Pro Glu His Phe Leu Asn Glu Asn Gly Lys Phe
385                 390                 395                 400

aag tac agt gac tat ttc aag cca ttt tcc aca gga aaa cga gtg tgt   1248
Lys Tyr Ser Asp Tyr Phe Lys Pro Phe Ser Thr Gly Lys Arg Val Cys
                405                 410                 415

gct gga gaa ggc ctg gct cgc atg gag ttg ttt ctt ttg ttg tgt gcc   1296
Ala Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Leu Leu Cys Ala
            420                 425                 430

att ttg cag cat ttt aat ttg aag cct ctc gtt gac cca aag gat atc   1344
Ile Leu Gln His Phe Asn Leu Lys Pro Leu Val Asp Pro Lys Asp Ile
        435                 440                 445

gac ctc agc cct ata cat att ggg ttt ggc tgt atc cca cca cgt tac   1392
Asp Leu Ser Pro Ile His Ile Gly Phe Gly Cys Ile Pro Pro Arg Tyr
    450                 455                 460

aaa ctc tgt gtc att ccc cgc tca tga                               1419
Lys Leu Cys Val Ile Pro Arg Ser
465                 470
```

<210> 4
<211> 472
<212> PRT
<213> Human cells

<400> 4

```
Met Ala Arg Gln Val His Ser Ser Trp Asn Leu Pro Pro Gly Pro Phe
 1               5                  10                  15

Pro Leu Pro Ile Ile Gly Asn Leu Phe Gln Leu Glu Leu Lys Asn Ile
            20                  25                  30

Pro Lys Ser Phe Thr Arg Leu Ala Gln Arg Phe Gly Pro Val Phe Thr
        35                  40                  45

Leu Tyr Val Gly Ser Gln Arg Met Val Val Met His Gly Tyr Lys Ala
    50                  55                  60

Val Lys Glu Ala Leu Leu Asp Tyr Lys Asp Glu Phe Ser Gly Arg Gly
65                  70                  75                  80

Asp Leu Pro Ala Phe His Ala His Arg Asp Arg Gly Ile Ile Phe Asn
                85                  90                  95

Asn Gly Pro Thr Trp Lys Asp Ile Arg Arg Phe Ser Leu Thr Thr Leu
            100                 105                 110

Arg Asn Tyr Gly Met Gly Lys Gln Gly Asn Glu Ser Arg Ile Gln Arg
        115                 120                 125

Glu Ala His Phe Leu Leu Glu Ala Leu Arg Lys Thr Gln Gly Gln Pro
    130                 135                 140

Phe Asp Pro Thr Phe Leu Ile Gly Cys Ala Pro Cys Asn Val Ile Ala
145                 150                 155                 160

Asp Ile Leu Phe Arg Lys His Phe Asp Tyr Asn Asp Glu Lys Phe Leu
                165                 170                 175

Arg Leu Met Tyr Leu Phe Asn Glu Asn Phe His Leu Leu Ser Thr Pro
            180                 185                 190

Trp Leu Gln Leu Tyr Asn Asn Phe Pro Ser Phe Leu His Tyr Leu Pro
        195                 200                 205
```

```
Gly Ser His Arg Lys Val Ile Lys Asn Val Ala Glu Val Lys Glu Tyr
    210                 215                 220

Val Ser Glu Arg Val Lys Glu His His Gln Ser Leu Asp Pro Asn Cys
225                 230                 235                 240

Pro Arg Asp Leu Thr Asp Cys Leu Leu Val Glu Met Glu Lys Glu Lys
                245                 250                 255

His Ser Ala Glu Arg Leu Tyr Thr Met Asp Gly Ile Thr Val Thr Val
            260                 265                 270

Ala Asp Leu Phe Phe Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu Arg
        275                 280                 285

Tyr Gly Leu Leu Ile Leu Met Lys Tyr Pro Glu Ile Glu Glu Lys Leu
    290                 295                 300

His Glu Glu Ile Asp Arg Val Ile Gly Pro Ser Arg Ile Pro Ala Ile
305                 310                 315                 320

Lys Asp Arg Gln Glu Met Pro Tyr Met Asp Ala Val Val His Glu Ile
                325                 330                 335

Gln Arg Phe Ile Thr Leu Val Pro Ser Asn Leu Pro His Glu Ala Thr
            340                 345                 350

Arg Asp Thr Ile Phe Arg Gly Tyr Leu Ile Pro Lys Gly Thr Val Val
        355                 360                 365

Val Pro Thr Leu Asp Ser Val Leu Tyr Asp Asn Gln Glu Phe Pro Asp
    370                 375                 380

Pro Glu Lys Phe Lys Pro Glu His Phe Leu Asn Glu Asn Gly Lys Phe
385                 390                 395                 400

Lys Tyr Ser Asp Tyr Phe Lys Pro Phe Ser Thr Gly Lys Arg Val Cys
                405                 410                 415
```

Ala Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Leu Leu Cys Ala
420 425 430

Ile Leu Gln His Phe Asn Leu Lys Pro Leu Val Asp Pro Lys Asp Ile
435 440 445

Asp Leu Ser Pro Ile His Ile Gly Phe Gly Cys Ile Pro Pro Arg Tyr
450 455 460

Lys Leu Cys Val Ile Pro Arg Ser
465 470

<210> 5
<211> 1438
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1416)

<400> 5
atg gct cga caa tct tct gga cga gga aaa ctc cct cct ggc cct act 48
Met Ala Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly Pro Thr
1 5 10 15

cct ctc cca gtg att gga aat atc cta cag ata gat att aag gat gtc 96
Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val
20 25 30

agc aaa tcc tta acc aat ctc tca aaa atc tat ggc cct gtg ttc act 144
Ser Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr
35 40 45

ctg tat ttt ggc ctg gaa cgc atg gtg gtg ctg cat gga tat gaa gtg 192
Leu Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val

```
     50                  55                  60

gtg aag gaa gcc ctg att gat ctt gga gag gag ttt tct gga aga ggc   240
Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly
 65                  70                  75                  80

cat ttc cca ctg gct gaa aga gct aac aga gga ttt gga atc gtt ttc   288
His Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe
                85                  90                  95

agc aat gga aag aga tgg aag gag atc cgg cgt ttc tcc ctc atg acg   336
Ser Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr
            100                 105                 110

ctg cgg aat ttt ggg atg ggg aag agg agc att gag gac cgt gtt caa   384
Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln
        115                 120                 125

gag gaa gcc cgc tgc ctt gtg gag gag ttg aga aaa acc aag gct tca   432
Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser
    130                 135                 140

ccc tgt gat ccc act ttc atc ctg ggc tgt gct ccc tgc aat gtg atc   480
Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile
145                 150                 155                 160

tgc tcc att att ttc cag aaa cgt ttc gat tat aaa gat cag caa ttt   528
Cys Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe
                165                 170                 175

ctt aac ttg atg gaa aaa ttg aat gaa aac atc agg att gta agc acc   576
Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr
            180                 185                 190

ccc tgg atc cag ata tgc aat aat ttt ccc act atc att gat tat ttc   624
Pro Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe
        195                 200                 205

ccg gga acc cat aac aaa tta ctt aaa aac ctt gct ttt atg gaa agt   672
```

```
Pro Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser
    210                 215                 220

gat att ttg gag aaa gta aaa gaa cac caa gaa tcg atg gac atc aac   720
Asp Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn
225                 230                 235                 240

aac cct cgg gac ttt att gat tgc ttc ctg atc aaa atg gag aag gaa   768
Asn Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu
                245                 250                 255

aag caa aac caa cag tct gaa ttc act att gaa aac ttg gta atc act   816
Lys Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr
            260                 265                 270

gca gct gac tta ctt gga gct ggg aca gag aca aca agc aca acc ctg   864
Ala Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
        275                 280                 285

aga tat gct ctc ctt ctc ctg ctg aag cac cca gag gtc aca gct aaa   912
Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
    290                 295                 300

gtc cag gaa gag att gaa cgt gtc gtt ggc aga aac cgg agc ccc tgc   960
Val Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys
305                 310                 315                 320

atg cag gac agg ggc cac atg ccc tac aca gat gct gtg gtg cac gag   1008
Met Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu
                325                 330                 335

gtc cag aga tac atc gac ctc atc ccc acc agc ctg ccc cat gca gtg   1056
Val Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val
            340                 345                 350

acc tgt gac gtt aaa ttc aga aac tac ctc att ccc aag ggc aca acc   1104
Thr Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
        355                 360                 365
```

```
ata tta act tcc ctc act tct gtg cta cat gac aac aaa gaa ttt ccc   1152
Ile Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
    370                 375                 380

aac cca gag atg ttt gac cct cgt cac ttt ctg gat gaa ggt gga aat   1200
Asn Pro Glu Met Phe Asp Pro Arg His Phe Leu Asp Glu Gly Gly Asn
385                 390                 395                 400

ttt aag aaa agt aac tac ttc atg cct ttc tca gca gga aaa cgg att   1248
Phe Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
                405                 410                 415

tgt gtg gga gag ggc ctg gcc cgc atg gag ctg ttt tta ttc ctg acc   1296
Cys Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr
            420                 425                 430

ttc att tta cag aac ttt aac ctg aaa tct ctg att gac cca aag gac   1344
Phe Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp
        435                 440                 445

ctt gac aca act cct gtt gtc aat gga ttt gct tct gtc ccg ccc ttc   1392
Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
    450                 455                 460

tat cag ctg tgc ttc att cct gtc tgaagaagca cagatggtct gg         1438
Tyr Gln Leu Cys Phe Ile Pro Val
465                 470
```

<210> 6
<211> 472
<212> PRT
<213> Human cells

<400> 6

```
Met Ala Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly Pro Thr
  1               5                  10                  15

Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val
```

```
                20                  25                  30

Ser Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr
         35                  40                  45

Leu Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val
     50                  55                  60

Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly
 65                  70                  75                  80

His Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe
                 85                  90                  95

Ser Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr
            100                 105                 110

Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln
        115                 120                 125

Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser
    130                 135                 140

Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile
145                 150                 155                 160

Cys Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe
                165                 170                 175

Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr
            180                 185                 190

Pro Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe
        195                 200                 205

Pro Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser
    210                 215                 220

Asp Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn
```

225 230 235 240

Asn Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu
245 250 255

Lys Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr
260 265 270

Ala Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
275 280 285

Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
290 295 300

Val Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys
305 310 315 320

Met Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu
325 330 335

Val Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val
340 345 350

Thr Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
355 360 365

Ile Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
370 375 380

Asn Pro Glu Met Phe Asp Pro Arg His Phe Leu Asp Glu Gly Gly Asn
385 390 395 400

Phe Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
405 410 415

Cys Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr
420 425 430

Phe Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp

```
        435                 440                 445

Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
    450                 455                 460

Tyr Gln Leu Cys Phe Ile Pro Val
465                 470
```

<210> 7
<211> 1485
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1482)

<400> 7

```
atg ctg gcc tca ggg atg ctt ctg gtg gcc ttg ctg gtc tgc ctg act   48
Met Leu Ala Ser Gly Met Leu Leu Val Ala Leu Leu Val Cys Leu Thr
  1               5                  10                  15

gtg atg gtc ttg atg tct gtt tgg cag cag agg aag agc aag ggg aag   96
Val Met Val Leu Met Ser Val Trp Gln Gln Arg Lys Ser Lys Gly Lys
             20                  25                  30

ctg cct ccg gga ccc acc cca ttg ccc ttc att gga aac tac ctg cag   144
Leu Pro Pro Gly Pro Thr Pro Leu Pro Phe Ile Gly Asn Tyr Leu Gln
         35                  40                  45

ctg aac aca gag cag atg tac aac tcc ctc atg aag atc agt gag cgc   192
Leu Asn Thr Glu Gln Met Tyr Asn Ser Leu Met Lys Ile Ser Glu Arg
     50                  55                  60

tat ggc ccc gtg ttc acc att cac ttg ggg ccc cgg cgg gtc gtg gtg   240
Tyr Gly Pro Val Phe Thr Ile His Leu Gly Pro Arg Arg Val Val Val
 65                  70                  75                  80
```

```
ctg tgt gga cat gat gcc gtc agg gag gct ctg gtg gac cag gct gag   288
Leu Cys Gly His Asp Ala Val Arg Glu Ala Leu Val Asp Gln Ala Glu
                85                  90                  95

gag ttc agc ggg cga ggc gag caa gcc acc ttc gac tgg gtc ttc aaa   336
Glu Phe Ser Gly Arg Gly Glu Gln Ala Thr Phe Asp Trp Val Phe Lys
            100                 105                 110

ggc tat ggc gtg gta ttc agc aac ggg gag cgc gcc aag cag ctc cgg   384
Gly Tyr Gly Val Val Phe Ser Asn Gly Glu Arg Ala Lys Gln Leu Arg
        115                 120                 125

cgc ttc tcc atc gcc acc ctg cgg gac ttc ggg gtg ggc aag cga ggc   432
Arg Phe Ser Ile Ala Thr Leu Arg Asp Phe Gly Val Gly Lys Arg Gly
    130                 135                 140

atc gag gag cgc atc cag gag gag gcg ggc ttc ctc atc gac gcc ctc   480
Ile Glu Glu Arg Ile Gln Glu Glu Ala Gly Phe Leu Ile Asp Ala Leu
145                 150                 155                 160

cgg ggc act ggc ggc gcc aat atc gat ccc acc ttc ttc ctg agc cgc   528
Arg Gly Thr Gly Gly Ala Asn Ile Asp Pro Thr Phe Phe Leu Ser Arg
                165                 170                 175

aca gtc tcc aat gtc atc agc tcc att gtc ttt ggg gac cgc ttt gac   576
Thr Val Ser Asn Val Ile Ser Ser Ile Val Phe Gly Asp Arg Phe Asp
            180                 185                 190

tat aag gac aaa gag ttc ctg tca ctg ttg cgc atg atg cta gga atc   624
Tyr Lys Asp Lys Glu Phe Leu Ser Leu Leu Arg Met Met Leu Gly Ile
        195                 200                 205

ttc cag ttc acg tca acc tcc acg ggg cag ctc tat gag atg ttc tct   672
Phe Gln Phe Thr Ser Thr Ser Thr Gly Gln Leu Tyr Glu Met Phe Ser
    210                 215                 220

tcg gtg atg aaa cac ctg cca gga cca cag caa cag gcc ttt cag ttg   720
Ser Val Met Lys His Leu Pro Gly Pro Gln Gln Gln Ala Phe Gln Leu
```

```
225                 230                 235                 240

ctg caa ggg ctg gag gac ttc ata gcc aag aag gtg gag cac aac cag   768
Leu Gln Gly Leu Glu Asp Phe Ile Ala Lys Lys Val Glu His Asn Gln
                245                 250                 255

cgc acg ctg gat ccc aat tcc cca cgg gac ttc att gac tcc ttt ctc   816
Arg Thr Leu Asp Pro Asn Ser Pro Arg Asp Phe Ile Asp Ser Phe Leu
            260                 265                 270

atc cgc atg cag gag gag gag aag aac ccc aac acg gag ttc tac ttg   864
Ile Arg Met Gln Glu Glu Glu Lys Asn Pro Asn Thr Glu Phe Tyr Leu
        275                 280                 285

aaa aac ctg gtg atg acc acg ttg aac ctc ttc att ggg ggc acc gag   912
Lys Asn Leu Val Met Thr Thr Leu Asn Leu Phe Ile Gly Gly Thr Glu
    290                 295                 300

acc gtc agc acc acc ctg cgc tat ggc ttc ttg ctg ctc atg aag cac   960
Thr Val Ser Thr Thr Leu Arg Tyr Gly Phe Leu Leu Leu Met Lys His
305                 310                 315                 320

cca gag gtg gag gcc aag gtc cat gag gag att gac aga gtg atc ggc   1008
Pro Glu Val Glu Ala Lys Val His Glu Glu Ile Asp Arg Val Ile Gly
                325                 330                 335

aag aac cgg cag ccc aag ttt gag gac cgg gcc aag atg ccc tac atg   1056
Lys Asn Arg Gln Pro Lys Phe Glu Asp Arg Ala Lys Met Pro Tyr Met
            340                 345                 350

gag gca gtg atc cac gag atc caa aga ttt gga gac gtg atc ccc atg   1104
Glu Ala Val Ile His Glu Ile Gln Arg Phe Gly Asp Val Ile Pro Met
        355                 360                 365

agt ttg gcc cgc aga gtc aaa aag gac acc aag ttt cgg gat ttc ttc   1152
Ser Leu Ala Arg Arg Val Lys Lys Asp Thr Lys Phe Arg Asp Phe Phe
    370                 375                 380

ctc cct aag ggc acc gaa gtg tac cct atg ctg ggc tct gtg ctg aga   1200
```

```
Leu Pro Lys Gly Thr Glu Val Tyr Pro Met Leu Gly Ser Val Leu Arg
385                 390                 395                 400

gac ccc agt ttc ttc tcc aac ccc cag gac ttc aat ccc cag cac ttc   1248
Asp Pro Ser Phe Phe Ser Asn Pro Gln Asp Phe Asn Pro Gln His Phe
                405                 410                 415

ctg aat gag aag ggg cag ttt aag aag agt gat gct ttt gtg ccc ttt   1296
Leu Asn Glu Lys Gly Gln Phe Lys Lys Ser Asp Ala Phe Val Pro Phe
            420                 425                 430

tcc atc gga aag cgg aac tgt ttc gga gaa ggc ctg gcc aga atg gag   1344
Ser Ile Gly Lys Arg Asn Cys Phe Gly Glu Gly Leu Ala Arg Met Glu
        435                 440                 445

ctc ttt ctc ttc ttc acc acc gtc atg cag aac ttc cgc ctc aag tcc   1392
Leu Phe Leu Phe Phe Thr Thr Val Met Gln Asn Phe Arg Leu Lys Ser
    450                 455                 460

tcc cag tca cct aag gac att gac gtg tcc ccc aaa cac gtg ggc ttt   1440
Ser Gln Ser Pro Lys Asp Ile Asp Val Ser Pro Lys His Val Gly Phe
465                 470                 475                 480

gcc acg atc cca cga aac tac acc atg agc ttc ctg ccc cgc tga       1485
Ala Thr Ile Pro Arg Asn Tyr Thr Met Ser Phe Leu Pro Arg
                485                 490
```

<210> 8
<211> 494
<212> PRT
<213> Human cells

<400> 8

```
Met Leu Ala Ser Gly Met Leu Leu Val Ala Leu Leu Val Cys Leu Thr
  1               5                  10                  15

Val Met Val Leu Met Ser Val Trp Gln Gln Arg Lys Ser Lys Gly Lys
            20                  25                  30
```

```
Leu Pro Pro Gly Pro Thr Pro Leu Pro Phe Ile Gly Asn Tyr Leu Gln
         35                  40                  45

Leu Asn Thr Glu Gln Met Tyr Asn Ser Leu Met Lys Ile Ser Glu Arg
     50                  55                  60

Tyr Gly Pro Val Phe Thr Ile His Leu Gly Pro Arg Arg Val Val Val
 65                  70                  75                  80

Leu Cys Gly His Asp Ala Val Arg Glu Ala Leu Val Asp Gln Ala Glu
                 85                  90                  95

Glu Phe Ser Gly Arg Gly Glu Gln Ala Thr Phe Asp Trp Val Phe Lys
            100                 105                 110

Gly Tyr Gly Val Val Phe Ser Asn Gly Glu Arg Ala Lys Gln Leu Arg
        115                 120                 125

Arg Phe Ser Ile Ala Thr Leu Arg Asp Phe Gly Val Gly Lys Arg Gly
    130                 135                 140

Ile Glu Glu Arg Ile Gln Glu Glu Ala Gly Phe Leu Ile Asp Ala Leu
145                 150                 155                 160

Arg Gly Thr Gly Gly Ala Asn Ile Asp Pro Thr Phe Phe Leu Ser Arg
                165                 170                 175

Thr Val Ser Asn Val Ile Ser Ser Ile Val Phe Gly Asp Arg Phe Asp
            180                 185                 190

Tyr Lys Asp Lys Glu Phe Leu Ser Leu Leu Arg Met Met Leu Gly Ile
        195                 200                 205

Phe Gln Phe Thr Ser Thr Ser Thr Gly Gln Leu Tyr Glu Met Phe Ser
    210                 215                 220

Ser Val Met Lys His Leu Pro Gly Pro Gln Gln Gln Ala Phe Gln Leu
225                 230                 235                 240
```

```
Leu Gln Gly Leu Glu Asp Phe Ile Ala Lys Lys Val Glu His Asn Gln
                245                 250                 255

Arg Thr Leu Asp Pro Asn Ser Pro Arg Asp Phe Ile Asp Ser Phe Leu
            260                 265                 270

Ile Arg Met Gln Glu Glu Glu Lys Asn Pro Asn Thr Glu Phe Tyr Leu
        275                 280                 285

Lys Asn Leu Val Met Thr Thr Leu Asn Leu Phe Ile Gly Gly Thr Glu
    290                 295                 300

Thr Val Ser Thr Thr Leu Arg Tyr Gly Phe Leu Leu Leu Met Lys His
305                 310                 315                 320

Pro Glu Val Glu Ala Lys Val His Glu Glu Ile Asp Arg Val Ile Gly
                325                 330                 335

Lys Asn Arg Gln Pro Lys Phe Glu Asp Arg Ala Lys Met Pro Tyr Met
            340                 345                 350

Glu Ala Val Ile His Glu Ile Gln Arg Phe Gly Asp Val Ile Pro Met
        355                 360                 365

Ser Leu Ala Arg Arg Val Lys Lys Asp Thr Lys Phe Arg Asp Phe Phe
    370                 375                 380

Leu Pro Lys Gly Thr Glu Val Tyr Pro Met Leu Gly Ser Val Leu Arg
385                 390                 395                 400

Asp Pro Ser Phe Phe Ser Asn Pro Gln Asp Phe Asn Pro Gln His Phe
                405                 410                 415

Leu Asn Glu Lys Gly Gln Phe Lys Lys Ser Asp Ala Phe Val Pro Phe
            420                 425                 430

Ser Ile Gly Lys Arg Asn Cys Phe Gly Glu Gly Leu Ala Arg Met Glu
        435                 440                 445
```

```
Leu Phe Leu Phe Phe Thr Thr Val Met Gln Asn Phe Arg Leu Lys Ser
    450                 455                 460

Ser Gln Ser Pro Lys Asp Ile Asp Val Ser Pro Lys His Val Gly Phe
465                 470                 475                 480

Ala Thr Ile Pro Arg Asn Tyr Thr Met Ser Phe Leu Pro Arg
                485                 490
```

<210> 9
<211> 1479
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1479)

<400> 9

```
atg tct gcc ctc gga gtc acc gtg gcc ctg ctg gtg tgg gcg gcc ttc   48
Met Ser Ala Leu Gly Val Thr Val Ala Leu Leu Val Trp Ala Ala Phe
  1               5                  10                  15

ctc ctg ctg gtg tcc atg tgg agg cag gtg cac agc agc tgg aat ctg   96
Leu Leu Leu Val Ser Met Trp Arg Gln Val His Ser Ser Trp Asn Leu
             20                  25                  30

ccc cca ggc cct ttc ccg ctt ccc atc atc ggg aac ctc ttc cag ttg   144
Pro Pro Gly Pro Phe Pro Leu Pro Ile Ile Gly Asn Leu Phe Gln Leu
         35                  40                  45

gaa ttg aag aat att ccc aag tcc ttc acc cgg ttg gcc cag cgc ttc   192
Glu Leu Lys Asn Ile Pro Lys Ser Phe Thr Arg Leu Ala Gln Arg Phe
     50                  55                  60

ggg ccg gtg ttc acg ctg tac gtg ggc tcg cag cgc atg gtg gtg atg   240
```

```
Gly Pro Val Phe Thr Leu Tyr Val Gly Ser Gln Arg Met Val Val Met
 65                  70                  75                  80

cac ggc tac aag gcg gtg aag gaa gcg ctg ctg gac tac aag gac gag   288
His Gly Tyr Lys Ala Val Lys Glu Ala Leu Leu Asp Tyr Lys Asp Glu
                 85                  90                  95

ttc tcg ggc aga ggc gac ctc ccc gcg ttc cat gcg cac agg gac agg   336
Phe Ser Gly Arg Gly Asp Leu Pro Ala Phe His Ala His Arg Asp Arg
            100                 105                 110

gga atc att ttt aat aat gga cct acc tgg aag gac atc cgg cgg ttt   384
Gly Ile Ile Phe Asn Asn Gly Pro Thr Trp Lys Asp Ile Arg Arg Phe
        115                 120                 125

tcc ctg acc acc ctc cgg aac tat ggg atg ggg aaa cag ggc aat gag   432
Ser Leu Thr Thr Leu Arg Asn Tyr Gly Met Gly Lys Gln Gly Asn Glu
    130                 135                 140

agc cgg atc cag agg gag gcc cac ttc ctg ctg gaa gca ctc agg aag   480
Ser Arg Ile Gln Arg Glu Ala His Phe Leu Leu Glu Ala Leu Arg Lys
145                 150                 155                 160

acc caa ggc cag cct ttc gac ccc acc ttc ctc atc ggc tgc gcg ccc   528
Thr Gln Gly Gln Pro Phe Asp Pro Thr Phe Leu Ile Gly Cys Ala Pro
                165                 170                 175

tgc aac gtc ata gcc gac atc ctc ttc cgc aag cat ttt gac tac aat   576
Cys Asn Val Ile Ala Asp Ile Leu Phe Arg Lys His Phe Asp Tyr Asn
            180                 185                 190

gat gag aag ttt cta agg ctg atg tat ttg ttt aat gag aac ttc cac   624
Asp Glu Lys Phe Leu Arg Leu Met Tyr Leu Phe Asn Glu Asn Phe His
        195                 200                 205

cta ctc agc act ccc tgg ctc cag ctt tac aat aat ttt ccc agc ttt   672
Leu Leu Ser Thr Pro Trp Leu Gln Leu Tyr Asn Asn Phe Pro Ser Phe
    210                 215                 220
```

```
cta cac tac ttg cct gga agc cac aga aaa gtc ata aaa aat gtg gct   720
Leu His Tyr Leu Pro Gly Ser His Arg Lys Val Ile Lys Asn Val Ala
225                 230                 235                 240

gaa gta aaa gag tat gtg tct gaa agg gtg aag gag cac cat caa tct   768
Glu Val Lys Glu Tyr Val Ser Glu Arg Val Lys Glu His His Gln Ser
                245                 250                 255

ctg gac ccc aac tgt ccc cgg gac ctc acc gac tgc ctg ctc gtg gaa   816
Leu Asp Pro Asn Cys Pro Arg Asp Leu Thr Asp Cys Leu Leu Val Glu
            260                 265                 270

atg gag aag gaa aag cac agt gca gag cgc ttg tac aca atg gac ggt   864
Met Glu Lys Glu Lys His Ser Ala Glu Arg Leu Tyr Thr Met Asp Gly
        275                 280                 285

atc acc gtg act gtg gcc gac ctg ttc ttt gcg ggg aca gag acc acc   912
Ile Thr Val Thr Val Ala Asp Leu Phe Phe Ala Gly Thr Glu Thr Thr
    290                 295                 300

agc aca act ctg aga tat ggg ctc ctg att ctc atg aaa tac cct gag   960
Ser Thr Thr Leu Arg Tyr Gly Leu Leu Ile Leu Met Lys Tyr Pro Glu
305                 310                 315                 320

atc gaa gag aag ctc cat gaa gaa att gac agg gtg att ggg cca agc   1008
Ile Glu Glu Lys Leu His Glu Glu Ile Asp Arg Val Ile Gly Pro Ser
                325                 330                 335

cga atc cct gcc atc aag gat agg caa gag atg ccc tac atg gat gct   1056
Arg Ile Pro Ala Ile Lys Asp Arg Gln Glu Met Pro Tyr Met Asp Ala
            340                 345                 350

gtg gtg cat gag att cag cgg ttc atc acc ctc gtg ccc tcc aac ctg   1104
Val Val His Glu Ile Gln Arg Phe Ile Thr Leu Val Pro Ser Asn Leu
        355                 360                 365

ccc cat gaa gca acc cga gac acc att ttc aga gga tac ctc atc ccc   1152
Pro His Glu Ala Thr Arg Asp Thr Ile Phe Arg Gly Tyr Leu Ile Pro
    370                 375                 380
```

```
aag ggc aca gtc gta gtg cca act ctg gac tct gtt ttg tat gac aac   1200
Lys Gly Thr Val Val Val Pro Thr Leu Asp Ser Val Leu Tyr Asp Asn
385                 390                 395                 400

caa gaa ttt cct gat cca gaa aag ttt aag cca gaa cac ttc ctg aat   1248
Gln Glu Phe Pro Asp Pro Glu Lys Phe Lys Pro Glu His Phe Leu Asn
                405                 410                 415

gaa aat gga aag ttc aag tac agt gac tat ttc aag cca ttt tcc aca   1296
Glu Asn Gly Lys Phe Lys Tyr Ser Asp Tyr Phe Lys Pro Phe Ser Thr
            420                 425                 430

gga aaa cga gtg tgt gct gga gaa ggc ctg gct cgc atg gag ttg ttt   1344
Gly Lys Arg Val Cys Ala Gly Glu Gly Leu Ala Arg Met Glu Leu Phe
        435                 440                 445

ctt ttg ttg tgt gcc att ttg cag cat ttt aat ttg aag cct ctc gtt   1392
Leu Leu Leu Cys Ala Ile Leu Gln His Phe Asn Leu Lys Pro Leu Val
    450                 455                 460

gac cca aag gat atc gac ctc agc cct ata cat att ggg ttt ggc tgt   1440
Asp Pro Lys Asp Ile Asp Leu Ser Pro Ile His Ile Gly Phe Gly Cys
465                 470                 475                 480

atc cca cca cgt tac aaa ctc tgt gtc att ccc cgc tca               1479
Ile Pro Pro Arg Tyr Lys Leu Cys Val Ile Pro Arg Ser
                485                 490
```

<210> 10
<211> 493
<212> PRT
<213> Human cells

<400> 10

```
Met Ser Ala Leu Gly Val Thr Val Ala Leu Leu Val Trp Ala Ala Phe
  1               5                  10                  15
```

```
Leu Leu Leu Val Ser Met Trp Arg Gln Val His Ser Ser Trp Asn Leu
            20                  25                  30

Pro Pro Gly Pro Phe Pro Leu Pro Ile Ile Gly Asn Leu Phe Gln Leu
        35                  40                  45

Glu Leu Lys Asn Ile Pro Lys Ser Phe Thr Arg Leu Ala Gln Arg Phe
    50                  55                  60

Gly Pro Val Phe Thr Leu Tyr Val Gly Ser Gln Arg Met Val Val Met
65                  70                  75                  80

His Gly Tyr Lys Ala Val Lys Glu Ala Leu Leu Asp Tyr Lys Asp Glu
                85                  90                  95

Phe Ser Gly Arg Gly Asp Leu Pro Ala Phe His Ala His Arg Asp Arg
            100                 105                 110

Gly Ile Ile Phe Asn Asn Gly Pro Thr Trp Lys Asp Ile Arg Arg Phe
        115                 120                 125

Ser Leu Thr Thr Leu Arg Asn Tyr Gly Met Gly Lys Gln Gly Asn Glu
    130                 135                 140

Ser Arg Ile Gln Arg Glu Ala His Phe Leu Leu Glu Ala Leu Arg Lys
145                 150                 155                 160

Thr Gln Gly Gln Pro Phe Asp Pro Thr Phe Leu Ile Gly Cys Ala Pro
                165                 170                 175

Cys Asn Val Ile Ala Asp Ile Leu Phe Arg Lys His Phe Asp Tyr Asn
            180                 185                 190

Asp Glu Lys Phe Leu Arg Leu Met Tyr Leu Phe Asn Glu Asn Phe His
        195                 200                 205

Leu Leu Ser Thr Pro Trp Leu Gln Leu Tyr Asn Asn Phe Pro Ser Phe
    210                 215                 220
```

```
Leu His Tyr Leu Pro Gly Ser His Arg Lys Val Ile Lys Asn Val Ala
225                 230                 235                 240

Glu Val Lys Glu Tyr Val Ser Glu Arg Val Lys Glu His His Gln Ser
                245                 250                 255

Leu Asp Pro Asn Cys Pro Arg Asp Leu Thr Asp Cys Leu Leu Val Glu
            260                 265                 270

Met Glu Lys Glu Lys His Ser Ala Glu Arg Leu Tyr Thr Met Asp Gly
        275                 280                 285

Ile Thr Val Thr Val Ala Asp Leu Phe Phe Ala Gly Thr Glu Thr Thr
    290                 295                 300

Ser Thr Thr Leu Arg Tyr Gly Leu Leu Ile Leu Met Lys Tyr Pro Glu
305                 310                 315                 320

Ile Glu Glu Lys Leu His Glu Glu Ile Asp Arg Val Ile Gly Pro Ser
                325                 330                 335

Arg Ile Pro Ala Ile Lys Asp Arg Gln Glu Met Pro Tyr Met Asp Ala
            340                 345                 350

Val Val His Glu Ile Gln Arg Phe Ile Thr Leu Val Pro Ser Asn Leu
        355                 360                 365

Pro His Glu Ala Thr Arg Asp Thr Ile Phe Arg Gly Tyr Leu Ile Pro
    370                 375                 380

Lys Gly Thr Val Val Val Pro Thr Leu Asp Ser Val Leu Tyr Asp Asn
385                 390                 395                 400

Gln Glu Phe Pro Asp Pro Glu Lys Phe Lys Pro Glu His Phe Leu Asn
                405                 410                 415

Glu Asn Gly Lys Phe Lys Tyr Ser Asp Tyr Phe Lys Pro Phe Ser Thr
            420                 425                 430
```

```
Gly Lys Arg Val Cys Ala Gly Glu Gly Leu Ala Arg Met Glu Leu Phe
        435                 440                 445

Leu Leu Leu Cys Ala Ile Leu Gln His Phe Asn Leu Lys Pro Leu Val
    450                 455                 460

Asp Pro Lys Asp Ile Asp Leu Ser Pro Ile His Ile Gly Phe Gly Cys
465                 470                 475                 480

Ile Pro Pro Arg Tyr Lys Leu Cys Val Ile Pro Arg Ser
                485                 490
```

<210> 11
<211> 1419
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1416)

<400> 11

```
atg gct aga cag agc tct ggg aga gga aaa ctc cct cct ggc cct act   48
Met Ala Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly Pro Thr
  1               5                  10                  15

cct ctc cca gtg att gga aat atc cta cag ata gat att aag gat gtc   96
Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val
             20                  25                  30

agc aaa tcc tta acc aat ctc tca aaa atc tat ggc cct gtg ttc act   144
Ser Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr
         35                  40                  45

ctg tat ttt ggc ctg gaa cgc atg gtg gtg ctg cat gga tat gaa gtg   192
Leu Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val
     50                  55                  60
```

```
gtg aag gaa gcc ctg att gat ctt gga gag gag ttt tct gga aga ggc   240
Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly
 65                  70                  75                  80

cat ttc cca ctg gct gaa aga gct aac aga gga ttt gga atc gtt ttc   288
His Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe
                 85                  90                  95

agc aat gga aag aga tgg aag gag atc cgg cgt ttc tcc ctc atg acg   336
Ser Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr
            100                 105                 110

ctg cgg aat ttt ggg atg ggg aag agg agc att gag gac cgt gtt caa   384
Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln
        115                 120                 125

gag gaa gcc cgc tgc ctt gtg gag gag ttg aga aaa acc aag gct tca   432
Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser
    130                 135                 140

ccc tgt gat ccc act ttc atc ctg ggc tgt gct ccc tgc aat gtg atc   480
Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile
145                 150                 155                 160

tgc tcc att att ttc cag aaa cgt ttc gat tat aaa gat cag caa ttt   528
Cys Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe
                165                 170                 175

ctt aac ttg atg gaa aaa ttg aat gaa aac atc agg att gta agc acc   576
Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr
            180                 185                 190

ccc tgg atc cag ata tgc aat aat ttt ccc act atc att gat tat ttc   624
Pro Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe
        195                 200                 205

ccg gga acc cat aac aaa tta ctt aaa aac ctt gct ttt atg gaa agt   672
Pro Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser
```

```
    210                 215                 220

gat att ttg gag aaa gta aaa gaa cac caa gaa tcg atg gac atc aac   720
Asp Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn
225                 230                 235                 240

aac cct cgg gac ttt att gat tgc ttc ctg atc aaa atg gag aag gaa   768
Asn Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu
                245                 250                 255

aag caa aac caa cag tct gaa ttc act att gaa aac ttg gta atc act   816
Lys Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr
            260                 265                 270

gca gct gac tta ctt gga gct ggg aca gag aca aca agc aca acc ctg   864
Ala Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
        275                 280                 285

aga tat gct ctc ctt ctc ctg ctg aag cac cca gag gtc aca gct aaa   912
Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
    290                 295                 300

gtc cag gaa gag att gaa cgt gtc gtt ggc aga aac cgg agc ccc tgc   960
Val Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys
305                 310                 315                 320

atg cag gac agg ggc cac atg ccc tac aca gat gct gtg gtg cac gag   1008
Met Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu
                325                 330                 335

gtc cag aga tac atc gac ctc atc ccc acc agc ctg ccc cat gca gtg   1056
Val Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val
            340                 345                 350

acc tgt gac gtt aaa ttc aga aac tac ctc att ccc aag ggc aca acc   1104
Thr Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
        355                 360                 365

ata tta act tcc ctc act tct gtg cta cat gac aac aaa gaa ttt ccc   1152
```

```
Ile Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
    370                 375                 380

aac cca gag atg ttt gac cct cgt cac ttt ctg gat gaa ggt gga aat   1200
Asn Pro Glu Met Phe Asp Pro Arg His Phe Leu Asp Glu Gly Gly Asn
385                 390                 395                 400

ttt aag aaa agt aac tac ttc atg cct ttc tca gca gga aaa cgg att   1248
Phe Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
                405                 410                 415

tgt gtg gga gag ggc ctg gcc cgc atg gag ctg ttt tta ttc ctg acc   1296
Cys Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr
            420                 425                 430

ttc att tta cag aac ttt aac ctg aaa tct ctg att gac cca aag gac   1344
Phe Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp
        435                 440                 445

ctt gac aca act cct gtt gtc aat gga ttt gct tct gtc ccg ccc ttc   1392
Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
    450                 455                 460

tat cag ctg tgc ttc att cct gtc tga                              1419
Tyr Gln Leu Cys Phe Ile Pro Val
465                 470
```

<210> 12
<211> 472
<212> PRT
<213> Human cells

```
<400> 12
Met Ala Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly Pro Thr
  1               5                  10                  15

Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val
            20                  25                  30
```

```
Ser Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr
         35                  40                  45

Leu Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val
     50                  55                  60

Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly
 65                  70                  75                  80

His Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe
                 85                  90                  95

Ser Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr
            100                 105                 110

Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln
        115                 120                 125

Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser
    130                 135                 140

Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile
145                 150                 155                 160

Cys Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe
                165                 170                 175

Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr
            180                 185                 190

Pro Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe
        195                 200                 205

Pro Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser
    210                 215                 220

Asp Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn
225                 230                 235                 240
```

```
Asn Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu
                245                 250                 255

Lys Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr
            260                 265                 270

Ala Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
        275                 280                 285

Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
    290                 295                 300

Val Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys
305                 310                 315                 320

Met Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu
                325                 330                 335

Val Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val
            340                 345                 350

Thr Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
        355                 360                 365

Ile Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
    370                 375                 380

Asn Pro Glu Met Phe Asp Pro Arg His Phe Leu Asp Glu Gly Gly Asn
385                 390                 395                 400

Phe Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
                405                 410                 415

Cys Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr
            420                 425                 430

Phe Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp
        435                 440                 445
```

```
Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
    450                 455                 460

Tyr Gln Leu Cys Phe Ile Pro Val
465                 470
```

<210> 13
<211> 2034
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(2031)

<400> 13

```
atg gct gac tcc cac gtg gac acc agc tcc acc gtg tcc gag gcg gtg   48
Met Ala Asp Ser His Val Asp Thr Ser Ser Thr Val Ser Glu Ala Val
  1               5                  10                  15

gcc gaa gaa gta tct ctt ttc agc atg acg gac atg att ctg ttt tcg   96
Ala Glu Glu Val Ser Leu Phe Ser Met Thr Asp Met Ile Leu Phe Ser
             20                  25                  30

ctc atc gtg ggt ctc cta acc tac tgg ttc ctc ttc aga aag aaa aaa   144
Leu Ile Val Gly Leu Leu Thr Tyr Trp Phe Leu Phe Arg Lys Lys Lys
         35                  40                  45

gaa gaa gtc ccc gag ttc acc aaa att cag aca ttg acc tcc tct gtc   192
Glu Glu Val Pro Glu Phe Thr Lys Ile Gln Thr Leu Thr Ser Ser Val
     50                  55                  60

aga gag agc agc ttt gtg gaa aag atg aag aaa acg ggg agg aac atc   240
Arg Glu Ser Ser Phe Val Glu Lys Met Lys Lys Thr Gly Arg Asn Ile
 65                  70                  75                  80
```

```
atc gtg ttc tac ggc tcc cag acg ggg act gca gag gag ttt gcc aac   288
Ile Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Phe Ala Asn
                85                  90                  95

cgc ctg tcc aag gac gcc cac cgc tac ggg atg cga ggc atg tca gcg   336
Arg Leu Ser Lys Asp Ala His Arg Tyr Gly Met Arg Gly Met Ser Ala
            100                 105                 110

gac cct gag gag tat gac ctg gcc gac ctg agc agc ctg cca gag atc   384
Asp Pro Glu Glu Tyr Asp Leu Ala Asp Leu Ser Ser Leu Pro Glu Ile
        115                 120                 125

gac aac gcc ctg gtg gtt ttc tgc atg gcc acc tac ggt gag gga gac   432
Asp Asn Ala Leu Val Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Asp
    130                 135                 140

ccc acc gac aat gcc cag gac ttc tac gac tgg ctg cag gag aca gac   480
Pro Thr Asp Asn Ala Gln Asp Phe Tyr Asp Trp Leu Gln Glu Thr Asp
145                 150                 155                 160

gtg gat ctc tct ggg gtc aag ttc gcg gtg ttt ggt ctt ggg aac aag   528
Val Asp Leu Ser Gly Val Lys Phe Ala Val Phe Gly Leu Gly Asn Lys
                165                 170                 175

acc tac gag cac ttc aat gcc atg ggc aag tac gtg gac aag cgg ctg   576
Thr Tyr Glu His Phe Asn Ala Met Gly Lys Tyr Val Asp Lys Arg Leu
            180                 185                 190

gag cag ctc ggc gcc cag cgc atc ttt gag ctg ggg ttg ggc gac gac   624
Glu Gln Leu Gly Ala Gln Arg Ile Phe Glu Leu Gly Leu Gly Asp Asp
        195                 200                 205

gat ggg aac ttg gag gag gac ttc atc acc tgg cga gag cag ttc tgg   672
Asp Gly Asn Leu Glu Glu Asp Phe Ile Thr Trp Arg Glu Gln Phe Trp
    210                 215                 220

ccg gcc gtg tgt gaa cac ttt ggg gtg gaa gcc act ggc gag gag tcc   720
Pro Ala Val Cys Glu His Phe Gly Val Glu Ala Thr Gly Glu Glu Ser
225                 230                 235                 240
```

```
agc att cgc cag tac gag ctt gtg gtc cac acc gac ata gat gcg gcc   768
Ser Ile Arg Gln Tyr Glu Leu Val Val His Thr Asp Ile Asp Ala Ala
                245                 250                 255

aag gtg tac atg ggg gag atg ggc cgg ctg aag agc tac gag aac cag   816
Lys Val Tyr Met Gly Glu Met Gly Arg Leu Lys Ser Tyr Glu Asn Gln
            260                 265                 270

aag ccc ccc ttt gat gcc aag aat ccg ttc ctg gct gca gtc acc acc   864
Lys Pro Pro Phe Asp Ala Lys Asn Pro Phe Leu Ala Ala Val Thr Thr
        275                 280                 285

aac cgg aag ctg aac cag gga acc gag cgc cac ctc atg cac ctg gaa   912
Asn Arg Lys Leu Asn Gln Gly Thr Glu Arg His Leu Met His Leu Glu
    290                 295                 300

ttg gac atc tcg gac tcc aaa atc agg tat gaa tct ggg gac cac gtg   960
Leu Asp Ile Ser Asp Ser Lys Ile Arg Tyr Glu Ser Gly Asp His Val
305                 310                 315                 320

gct gtg tac cca gcc aac gac tct gct ctc gtc aac cag ctg ggc aaa   1008
Ala Val Tyr Pro Ala Asn Asp Ser Ala Leu Val Asn Gln Leu Gly Lys
                325                 330                 335

atc ctg ggt gcc gac ctg gac gtc gtc atg tcc ctg aac aac ctg gat   1056
Ile Leu Gly Ala Asp Leu Asp Val Val Met Ser Leu Asn Asn Leu Asp
            340                 345                 350

gag gag tcc aac aag aag cac cca ttc ccg tgc cct acg tcc tac cgc   1104
Glu Glu Ser Asn Lys Lys His Pro Phe Pro Cys Pro Thr Ser Tyr Arg
        355                 360                 365

acg gcc ctc acc tac tac ctg gac atc acc aac ccg ccg cgt acc aac   1152
Thr Ala Leu Thr Tyr Tyr Leu Asp Ile Thr Asn Pro Pro Arg Thr Asn
    370                 375                 380

gtg ctg tac gag ctg gcg cag tac gcc tcg gag ccc tcg gag cag gag   1200
Val Leu Tyr Glu Leu Ala Gln Tyr Ala Ser Glu Pro Ser Glu Gln Glu
```

```
385                 390                 395                 400

ctg ctg cgc aag atg gcc tcc tcc tcc ggc gag ggc aag gag ctg tac   1248
Leu Leu Arg Lys Met Ala Ser Ser Ser Gly Glu Gly Lys Glu Leu Tyr
                405                 410                 415

ctg agc tgg gtg gtg gag gcc cgg agg cac atc ctg gcc atc ctg cag   1296
Leu Ser Trp Val Val Glu Ala Arg Arg His Ile Leu Ala Ile Leu Gln
            420                 425                 430

gac tgc ccg tcc ctg cgg ccc ccc atc gac cac ctg tgt gag ctg ctg   1344
Asp Cys Pro Ser Leu Arg Pro Pro Ile Asp His Leu Cys Glu Leu Leu
        435                 440                 445

ccg cgc ctg cag gcc cgc tac tac tcc atc gcc tca tcc tcc aag gtc   1392
Pro Arg Leu Gln Ala Arg Tyr Tyr Ser Ile Ala Ser Ser Ser Lys Val
    450                 455                 460

cac ccc aac tct gtg cac atc tgt gcg gtg gtt gtg gag tac gag acc   1440
His Pro Asn Ser Val His Ile Cys Ala Val Val Val Glu Tyr Glu Thr
465                 470                 475                 480

aag gcc ggc cgc atc aac aag ggc gtg gcc acc aac tgg ctg cgg gcc   1488
Lys Ala Gly Arg Ile Asn Lys Gly Val Ala Thr Asn Trp Leu Arg Ala
                485                 490                 495

aag gag cct gcc ggg gag aac ggc ggc cgt gcg ctg gtg ccc atg ttc   1536
Lys Glu Pro Ala Gly Glu Asn Gly Gly Arg Ala Leu Val Pro Met Phe
            500                 505                 510

gtg cgc aag tcc cag ttc cgc ctg ccc ttc aag gcc acc acg cct gtc   1584
Val Arg Lys Ser Gln Phe Arg Leu Pro Phe Lys Ala Thr Thr Pro Val
        515                 520                 525

atc atg gtg ggc ccc ggc acc ggg gtg gca ccc ttc ata ggc ttc atc   1632
Ile Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Ile Gly Phe Ile
    530                 535                 540

cag gag cgg gcc tgg ctg cga cag cag ggc aag gag gtg ggg gag acg   1680
```

```
Gln Glu Arg Ala Trp Leu Arg Gln Gln Gly Lys Glu Val Gly Glu Thr
545                 550                 555                 560

ctg ctg tac tac ggc tgc cgc cgc tcg gat gag gac tac ctg tac cgg   1728
Leu Leu Tyr Tyr Gly Cys Arg Arg Ser Asp Glu Asp Tyr Leu Tyr Arg
                565                 570                 575

gag gac gtg gcg cag ttc cac agg gac ggt gcg ctc acc cag ctc aac   1776
Glu Asp Val Ala Gln Phe His Arg Asp Gly Ala Leu Thr Gln Leu Asn
            580                 585                 590

gtg gcc ttc tcc cgg gag cag tcc cac aag gtc tac gtc cag cac ctg   1824
Val Ala Phe Ser Arg Glu Gln Ser His Lys Val Tyr Val Gln His Leu
        595                 600                 605

cta aag caa gac cga gag cac ctg tgg aag ttg atc gaa ggc ggt gcc   1872
Leu Lys Gln Asp Arg Glu His Leu Trp Lys Leu Ile Glu Gly Gly Ala
    610                 615                 620

cac atc tac gtc tgt ggg gat gca cgg aac atg gcc agg gat gtg cag   1920
His Ile Tyr Val Cys Gly Asp Ala Arg Asn Met Ala Arg Asp Val Gln
625                 630                 635                 640

aac acc ttc tac gac atc gtg gct gag ctc ggg gcc atg gag cac gcg   1968
Asn Thr Phe Tyr Asp Ile Val Ala Glu Leu Gly Ala Met Glu His Ala
                645                 650                 655

cag gcg gtg gac tac atc aag aaa ctg atg acc aag ggc cgc tac tcc   2016
Gln Ala Val Asp Tyr Ile Lys Lys Leu Met Thr Lys Gly Arg Tyr Ser
            660                 665                 670

ctg gac gtg tgg agc tag                                          2034
Leu Asp Val Trp Ser
        675
```

<210> 14
<211> 677
<212> PRT

<213> Human cells

<400> 14

```
Met Ala Asp Ser His Val Asp Thr Ser Ser Thr Val Ser Glu Ala Val
 1               5                  10                  15

Ala Glu Glu Val Ser Leu Phe Ser Met Thr Asp Met Ile Leu Phe Ser
            20                  25                  30

Leu Ile Val Gly Leu Leu Thr Tyr Trp Phe Leu Phe Arg Lys Lys Lys
        35                  40                  45

Glu Glu Val Pro Glu Phe Thr Lys Ile Gln Thr Leu Thr Ser Ser Val
    50                  55                  60

Arg Glu Ser Ser Phe Val Glu Lys Met Lys Lys Thr Gly Arg Asn Ile
65                  70                  75                  80

Ile Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Phe Ala Asn
                85                  90                  95

Arg Leu Ser Lys Asp Ala His Arg Tyr Gly Met Arg Gly Met Ser Ala
            100                 105                 110

Asp Pro Glu Glu Tyr Asp Leu Ala Asp Leu Ser Ser Leu Pro Glu Ile
        115                 120                 125

Asp Asn Ala Leu Val Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Asp
    130                 135                 140

Pro Thr Asp Asn Ala Gln Asp Phe Tyr Asp Trp Leu Gln Glu Thr Asp
145                 150                 155                 160

Val Asp Leu Ser Gly Val Lys Phe Ala Val Phe Gly Leu Gly Asn Lys
                165                 170                 175

Thr Tyr Glu His Phe Asn Ala Met Gly Lys Tyr Val Asp Lys Arg Leu
            180                 185                 190
```

```
Glu Gln Leu Gly Ala Gln Arg Ile Phe Glu Leu Gly Leu Gly Asp Asp
        195                 200                 205

Asp Gly Asn Leu Glu Glu Asp Phe Ile Thr Trp Arg Glu Gln Phe Trp
    210                 215                 220

Pro Ala Val Cys Glu His Phe Gly Val Glu Ala Thr Gly Glu Glu Ser
225                 230                 235                 240

Ser Ile Arg Gln Tyr Glu Leu Val Val His Thr Asp Ile Asp Ala Ala
                245                 250                 255

Lys Val Tyr Met Gly Glu Met Gly Arg Leu Lys Ser Tyr Glu Asn Gln
            260                 265                 270

Lys Pro Pro Phe Asp Ala Lys Asn Pro Phe Leu Ala Ala Val Thr Thr
        275                 280                 285

Asn Arg Lys Leu Asn Gln Gly Thr Glu Arg His Leu Met His Leu Glu
    290                 295                 300

Leu Asp Ile Ser Asp Ser Lys Ile Arg Tyr Glu Ser Gly Asp His Val
305                 310                 315                 320

Ala Val Tyr Pro Ala Asn Asp Ser Ala Leu Val Asn Gln Leu Gly Lys
                325                 330                 335

Ile Leu Gly Ala Asp Leu Asp Val Val Met Ser Leu Asn Asn Leu Asp
            340                 345                 350

Glu Glu Ser Asn Lys Lys His Pro Phe Pro Cys Pro Thr Ser Tyr Arg
        355                 360                 365

Thr Ala Leu Thr Tyr Tyr Leu Asp Ile Thr Asn Pro Pro Arg Thr Asn
    370                 375                 380

Val Leu Tyr Glu Leu Ala Gln Tyr Ala Ser Glu Pro Ser Glu Gln Glu
385                 390                 395                 400
```

```
Leu Leu Arg Lys Met Ala Ser Ser Ser Gly Glu Gly Lys Glu Leu Tyr
                405                 410                 415

Leu Ser Trp Val Val Glu Ala Arg Arg His Ile Leu Ala Ile Leu Gln
            420                 425                 430

Asp Cys Pro Ser Leu Arg Pro Pro Ile Asp His Leu Cys Glu Leu Leu
        435                 440                 445

Pro Arg Leu Gln Ala Arg Tyr Tyr Ser Ile Ala Ser Ser Ser Lys Val
    450                 455                 460

His Pro Asn Ser Val His Ile Cys Ala Val Val Val Glu Tyr Glu Thr
465                 470                 475                 480

Lys Ala Gly Arg Ile Asn Lys Gly Val Ala Thr Asn Trp Leu Arg Ala
                485                 490                 495

Lys Glu Pro Ala Gly Glu Asn Gly Gly Arg Ala Leu Val Pro Met Phe
            500                 505                 510

Val Arg Lys Ser Gln Phe Arg Leu Pro Phe Lys Ala Thr Thr Pro Val
        515                 520                 525

Ile Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Ile Gly Phe Ile
    530                 535                 540

Gln Glu Arg Ala Trp Leu Arg Gln Gln Gly Lys Glu Val Gly Glu Thr
545                 550                 555                 560

Leu Leu Tyr Tyr Gly Cys Arg Arg Ser Asp Glu Asp Tyr Leu Tyr Arg
                565                 570                 575

Glu Asp Val Ala Gln Phe His Arg Asp Gly Ala Leu Thr Gln Leu Asn
            580                 585                 590

Val Ala Phe Ser Arg Glu Gln Ser His Lys Val Tyr Val Gln His Leu
        595                 600                 605
```

```
Leu Lys Gln Asp Arg Glu His Leu Trp Lys Leu Ile Glu Gly Gly Ala
    610                 615                 620

His Ile Tyr Val Cys Gly Asp Ala Arg Asn Met Ala Arg Asp Val Gln
625                 630                 635                 640

Asn Thr Phe Tyr Asp Ile Val Ala Glu Leu Gly Ala Met Glu His Ala
                645                 650                 655

Gln Ala Val Asp Tyr Ile Lys Lys Leu Met Thr Lys Gly Arg Tyr Ser
            660                 665                 670

Leu Asp Val Trp Ser
        675
```

<210> 15
<211> 2034
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(2031)

<400> 15

```
atg gga gac tcc cac gtg gac acc agc tcc acc gtg tcc gag gcg gtg   48
Met Gly Asp Ser His Val Asp Thr Ser Ser Thr Val Ser Glu Ala Val
  1               5                  10                  15

gcc gaa gaa gta tct ctt ttc agc atg acg gac atg att ctg ttt tcg   96
Ala Glu Glu Val Ser Leu Phe Ser Met Thr Asp Met Ile Leu Phe Ser
             20                  25                  30

ctc atc gtg ggt ctc cta acc tac tgg ttc ctc ttc aga aag aaa aaa   144
Leu Ile Val Gly Leu Leu Thr Tyr Trp Phe Leu Phe Arg Lys Lys Lys
         35                  40                  45
```

```
gaa gaa gtc ccc gag ttc acc aaa att cag aca ttg acc tcc tct gtc   192
Glu Glu Val Pro Glu Phe Thr Lys Ile Gln Thr Leu Thr Ser Ser Val
     50                  55                  60

aga gag agc agc ttt gtg gaa aag atg aag aaa acg ggg agg aac atc   240
Arg Glu Ser Ser Phe Val Glu Lys Met Lys Lys Thr Gly Arg Asn Ile
 65                  70                  75                  80

atc gtg ttc tac ggc tcc cag acg ggg act gca gag gag ttt gcc aac   288
Ile Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Phe Ala Asn
                 85                  90                  95

cgc ctg tcc aag gac gcc cac cgc tac ggg atg cga ggc atg tca gcg   336
Arg Leu Ser Lys Asp Ala His Arg Tyr Gly Met Arg Gly Met Ser Ala
            100                 105                 110

gac cct gag gag tat gac ctg gcc gac ctg agc agc ctg cca gag atc   384
Asp Pro Glu Glu Tyr Asp Leu Ala Asp Leu Ser Ser Leu Pro Glu Ile
        115                 120                 125

gac aac gcc ctg gtg gtt ttc tgc atg gcc acc tac ggt gag gga gac   432
Asp Asn Ala Leu Val Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Asp
    130                 135                 140

ccc acc gac aat gcc cag gac ttc tac gac tgg ctg cag gag aca gac   480
Pro Thr Asp Asn Ala Gln Asp Phe Tyr Asp Trp Leu Gln Glu Thr Asp
145                 150                 155                 160

gtg gat ctc tct ggg gtc aag ttc gcg gtg ttt ggt ctt ggg aac aag   528
Val Asp Leu Ser Gly Val Lys Phe Ala Val Phe Gly Leu Gly Asn Lys
                165                 170                 175

acc tac gag cac ttc aat gcc atg ggc aag tac gtg gac aag cgg ctg   576
Thr Tyr Glu His Phe Asn Ala Met Gly Lys Tyr Val Asp Lys Arg Leu
            180                 185                 190

gag cag ctc ggc gcc cag cgc atc ttt gag ctg ggg ttg ggc gac gac   624
Glu Gln Leu Gly Ala Gln Arg Ile Phe Glu Leu Gly Leu Gly Asp Asp
        195                 200                 205
```

```
gat ggg aac ttg gag gag gac ttc atc acc tgg cga gag cag ttc tgg   672
Asp Gly Asn Leu Glu Glu Asp Phe Ile Thr Trp Arg Glu Gln Phe Trp
    210                 215                 220

ccg gcc gtg tgt gaa cac ttt ggg gtg gaa gcc act ggc gag gag tcc   720
Pro Ala Val Cys Glu His Phe Gly Val Glu Ala Thr Gly Glu Glu Ser
225                 230                 235                 240

agc att cgc cag tac gag ctt gtg gtc cac acc gac ata gat gcg gcc   768
Ser Ile Arg Gln Tyr Glu Leu Val Val His Thr Asp Ile Asp Ala Ala
                245                 250                 255

aag gtg tac atg ggg gag atg ggc cgg ctg aag agc tac gag aac cag   816
Lys Val Tyr Met Gly Glu Met Gly Arg Leu Lys Ser Tyr Glu Asn Gln
            260                 265                 270

aag ccc ccc ttt gat gcc aag aat ccg ttc ctg gct gca gtc acc acc   864
Lys Pro Pro Phe Asp Ala Lys Asn Pro Phe Leu Ala Ala Val Thr Thr
        275                 280                 285

aac cgg aag ctg aac cag gga acc gag cgc cac ctc atg cac ctg gaa   912
Asn Arg Lys Leu Asn Gln Gly Thr Glu Arg His Leu Met His Leu Glu
    290                 295                 300

ttg gac atc tcg gac tcc aaa atc agg tat gaa tct ggg gac cac gtg   960
Leu Asp Ile Ser Asp Ser Lys Ile Arg Tyr Glu Ser Gly Asp His Val
305                 310                 315                 320

gct gtg tac cca gcc aac gac tct gct ctc gtc aac cag ctg ggc aaa   1008
Ala Val Tyr Pro Ala Asn Asp Ser Ala Leu Val Asn Gln Leu Gly Lys
                325                 330                 335

atc ctg ggt gcc gac ctg gac gtc gtc atg tcc ctg aac aac ctg gat   1056
Ile Leu Gly Ala Asp Leu Asp Val Val Met Ser Leu Asn Asn Leu Asp
            340                 345                 350

gag gag tcc aac aag aag cac cca ttc ccg tgc cct acg tcc tac cgc   1104
Glu Glu Ser Asn Lys Lys His Pro Phe Pro Cys Pro Thr Ser Tyr Arg
```

```
        355                 360                 365

acg gcc ctc acc tac tac ctg gac atc acc aac ccg ccg cgt acc aac   1152
Thr Ala Leu Thr Tyr Tyr Leu Asp Ile Thr Asn Pro Pro Arg Thr Asn
    370                 375                 380

gtg ctg tac gag ctg gcg cag tac gcc tcg gag ccc tcg gag cag gag   1200
Val Leu Tyr Glu Leu Ala Gln Tyr Ala Ser Glu Pro Ser Glu Gln Glu
385                 390                 395                 400

ctg ctg cgc aag atg gcc tcc tcc tcc ggc gag ggc aag gag ctg tac   1248
Leu Leu Arg Lys Met Ala Ser Ser Ser Gly Glu Gly Lys Glu Leu Tyr
                405                 410                 415

ctg agc tgg gtg gtg gag gcc cgg agg cac atc ctg gcc atc ctg cag   1296
Leu Ser Trp Val Val Glu Ala Arg Arg His Ile Leu Ala Ile Leu Gln
            420                 425                 430

gac tgc ccg tcc ctg cgg ccc ccc atc gac cac ctg tgt gag ctg ctg   1344
Asp Cys Pro Ser Leu Arg Pro Pro Ile Asp His Leu Cys Glu Leu Leu
        435                 440                 445

ccg cgc ctg cag gcc cgc tac tac tcc atc gcc tca tcc tcc aag gtc   1392
Pro Arg Leu Gln Ala Arg Tyr Tyr Ser Ile Ala Ser Ser Ser Lys Val
    450                 455                 460

cac ccc aac tct gtg cac atc tgt gcg gtg gtt gtg gag tac gag acc   1440
His Pro Asn Ser Val His Ile Cys Ala Val Val Val Glu Tyr Glu Thr
465                 470                 475                 480

aag gcc ggc cgc atc aac aag ggc gtg gcc acc aac tgg ctg cgg gcc   1488
Lys Ala Gly Arg Ile Asn Lys Gly Val Ala Thr Asn Trp Leu Arg Ala
                485                 490                 495

aag gag cct gcc ggg gag aac ggc ggc cgt gcg ctg gtg ccc atg ttc   1536
Lys Glu Pro Ala Gly Glu Asn Gly Gly Arg Ala Leu Val Pro Met Phe
            500                 505                 510

gtg cgc aag tcc cag ttc cgc ctg ccc ttc aag gcc acc acg cct gtc   1584
```

```
Val Arg Lys Ser Gln Phe Arg Leu Pro Phe Lys Ala Thr Thr Pro Val
        515                 520                 525

atc atg gtg ggc ccc ggc acc ggg gtg gca ccc ttc ata ggc ttc atc   1632
Ile Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Ile Gly Phe Ile
    530                 535                 540

cag gag cgg gcc tgg ctg cga cag cag ggc aag gag gtg ggg gag acg   1680
Gln Glu Arg Ala Trp Leu Arg Gln Gln Gly Lys Glu Val Gly Glu Thr
545                 550                 555                 560

ctg ctg tac tac ggc tgc cgc cgc tcg gat gag gac tac ctg tac cgg   1728
Leu Leu Tyr Tyr Gly Cys Arg Arg Ser Asp Glu Asp Tyr Leu Tyr Arg
                565                 570                 575

gag gac gtg gcg cag ttc cac agg gac ggt gcg ctc acc cag ctc aac   1776
Glu Asp Val Ala Gln Phe His Arg Asp Gly Ala Leu Thr Gln Leu Asn
            580                 585                 590

gtg gcc ttc tcc cgg gag cag tcc cac aag gtc tac gtc cag cac ctg   1824
Val Ala Phe Ser Arg Glu Gln Ser His Lys Val Tyr Val Gln His Leu
        595                 600                 605

cta aag caa gac cga gag cac ctg tgg aag ttg atc gaa ggc ggt gcc   1872
Leu Lys Gln Asp Arg Glu His Leu Trp Lys Leu Ile Glu Gly Gly Ala
    610                 615                 620

cac atc tac gtc tgt ggg gat gca cgg aac atg gcc agg gat gtg cag   1920
His Ile Tyr Val Cys Gly Asp Ala Arg Asn Met Ala Arg Asp Val Gln
625                 630                 635                 640

aac acc ttc tac gac atc gtg gct gag ctc ggg gcc atg gag cac gcg   1968
Asn Thr Phe Tyr Asp Ile Val Ala Glu Leu Gly Ala Met Glu His Ala
                645                 650                 655

cag gcg gtg gac tac atc aag aaa ctg atg acc aag ggc cgc tac tcc   2016
Gln Ala Val Asp Tyr Ile Lys Lys Leu Met Thr Lys Gly Arg Tyr Ser
            660                 665                 670
```

ctg gac gtg tgg agc tag 2034
Leu Asp Val Trp Ser
675

<210> 16
<211> 677
<212> PRT
<213> Human cells

<400> 16
Met Gly Asp Ser His Val Asp Thr Ser Ser Thr Val Ser Glu Ala Val
1 5 10 15

Ala Glu Glu Val Ser Leu Phe Ser Met Thr Asp Met Ile Leu Phe Ser
20 25 30

Leu Ile Val Gly Leu Leu Thr Tyr Trp Phe Leu Phe Arg Lys Lys Lys
35 40 45

Glu Glu Val Pro Glu Phe Thr Lys Ile Gln Thr Leu Thr Ser Ser Val
50 55 60

Arg Glu Ser Ser Phe Val Glu Lys Met Lys Lys Thr Gly Arg Asn Ile
65 70 75 80

Ile Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Phe Ala Asn
85 90 95

Arg Leu Ser Lys Asp Ala His Arg Tyr Gly Met Arg Gly Met Ser Ala
100 105 110

Asp Pro Glu Glu Tyr Asp Leu Ala Asp Leu Ser Ser Leu Pro Glu Ile
115 120 125

Asp Asn Ala Leu Val Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Asp
130 135 140

Pro Thr Asp Asn Ala Gln Asp Phe Tyr Asp Trp Leu Gln Glu Thr Asp

145 150 155 160

Val Asp Leu Ser Gly Val Lys Phe Ala Val Phe Gly Leu Gly Asn Lys
165 170 175

Thr Tyr Glu His Phe Asn Ala Met Gly Lys Tyr Val Asp Lys Arg Leu
180 185 190

Glu Gln Leu Gly Ala Gln Arg Ile Phe Glu Leu Gly Leu Gly Asp Asp
195 200 205

Asp Gly Asn Leu Glu Glu Asp Phe Ile Thr Trp Arg Glu Gln Phe Trp
210 215 220

Pro Ala Val Cys Glu His Phe Gly Val Glu Ala Thr Gly Glu Glu Ser
225 230 235 240

Ser Ile Arg Gln Tyr Glu Leu Val Val His Thr Asp Ile Asp Ala Ala
245 250 255

Lys Val Tyr Met Gly Glu Met Gly Arg Leu Lys Ser Tyr Glu Asn Gln
260 265 270

Lys Pro Pro Phe Asp Ala Lys Asn Pro Phe Leu Ala Ala Val Thr Thr
275 280 285

Asn Arg Lys Leu Asn Gln Gly Thr Glu Arg His Leu Met His Leu Glu
290 295 300

Leu Asp Ile Ser Asp Ser Lys Ile Arg Tyr Glu Ser Gly Asp His Val
305 310 315 320

Ala Val Tyr Pro Ala Asn Asp Ser Ala Leu Val Asn Gln Leu Gly Lys
325 330 335

Ile Leu Gly Ala Asp Leu Asp Val Val Met Ser Leu Asn Asn Leu Asp
340 345 350

Glu Glu Ser Asn Lys Lys His Pro Phe Pro Cys Pro Thr Ser Tyr Arg

355 360 365

Thr Ala Leu Thr Tyr Tyr Leu Asp Ile Thr Asn Pro Pro Arg Thr Asn
370 375 380

Val Leu Tyr Glu Leu Ala Gln Tyr Ala Ser Glu Pro Ser Glu Gln Glu
385 390 395 400

Leu Leu Arg Lys Met Ala Ser Ser Ser Gly Glu Gly Lys Glu Leu Tyr
405 410 415

Leu Ser Trp Val Val Glu Ala Arg Arg His Ile Leu Ala Ile Leu Gln
420 425 430

Asp Cys Pro Ser Leu Arg Pro Pro Ile Asp His Leu Cys Glu Leu Leu
435 440 445

Pro Arg Leu Gln Ala Arg Tyr Tyr Ser Ile Ala Ser Ser Ser Lys Val
450 455 460

His Pro Asn Ser Val His Ile Cys Ala Val Val Val Glu Tyr Glu Thr
465 470 475 480

Lys Ala Gly Arg Ile Asn Lys Gly Val Ala Thr Asn Trp Leu Arg Ala
485 490 495

Lys Glu Pro Ala Gly Glu Asn Gly Gly Arg Ala Leu Val Pro Met Phe
500 505 510

Val Arg Lys Ser Gln Phe Arg Leu Pro Phe Lys Ala Thr Thr Pro Val
515 520 525

Ile Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Ile Gly Phe Ile
530 535 540

Gln Glu Arg Ala Trp Leu Arg Gln Gln Gly Lys Glu Val Gly Glu Thr
545 550 555 560

Leu Leu Tyr Tyr Gly Cys Arg Arg Ser Asp Glu Asp Tyr Leu Tyr Arg

565 570 575

Glu Asp Val Ala Gln Phe His Arg Asp Gly Ala Leu Thr Gln Leu Asn
580 585 590

Val Ala Phe Ser Arg Glu Gln Ser His Lys Val Tyr Val Gln His Leu
595 600 605

Leu Lys Gln Asp Arg Glu His Leu Trp Lys Leu Ile Glu Gly Gly Ala
610 615 620

His Ile Tyr Val Cys Gly Asp Ala Arg Asn Met Ala Arg Asp Val Gln
625 630 635 640

Asn Thr Phe Tyr Asp Ile Val Ala Glu Leu Gly Ala Met Glu His Ala
645 650 655

Gln Ala Val Asp Tyr Ile Lys Lys Leu Met Thr Lys Gly Arg Tyr Ser
660 665 670

Leu Asp Val Trp Ser
675

<210> 17
<211> 8537
<212> DNA
<213> Human cells

<400> 17
atggctgctt caggaatgtt attagtggcc ttgctggtct gcctgactgt gatggtcttg 60
atgtctgttt ggcagcagag gaagagcaag gggaagctgc ctccgggacc caccccattg 120
cccttcattg gaaactacct gcagctgaac acagagcaga tgtacaactc cctcatgaag 180
atcagtgagc gctatggccc cgtgttcacc attcacttgg ggccccggcg ggtcgtggtg 240
ctgtgtggac atgatgccgt cagggaggct ctggtggacc aggctgagga gttcagcggg 300
cgaggcgagc aagccacctt cgactgggtc ttcaaaggct atggcgtggt attcagcaac 360
ggggagcgcg ccaagcagct ccggcgcttc tccatcgcca ccctgcggga cttcggggtg 420
ggcaagcgag gcatcgagga gcgcatccag gaggaggcgg gcttcctcat cgacgccctc 480

cggggcactg gcggcgccaa tatcgatccc accttcttcc tgagccgcac agtctccaat 540
gtcatcagct ccattgtctt tggggaccgc tttgactata aggacaaaga gttcctgtca 600
ctgttgcgca tgatgctagg aatcttccag ttcacgtcaa cctccacggg gcagctctat 660
gagatgttct cttcggtgat gaaacacctg ccaggaccac agcaacaggc ctttcagttg 720
ctgcaagggc tggaggactt catagccaag aaggtggagc acaaccagcg cacgctggat 780
cccaattccc cacgggactt cattgactcc tttctcatcc gcatgcagga ggaggagaag 840
aaccccaaca cggagttcta cttgaaaaac ctggtgatga ccacgttgaa cctcttcatt 900
gggggcaccg agaccgtcag caccaccctg cgctatggct tcttgctgct catgaagcac 960
ccagaggtgg aggccaaggt ccatgaggag attgacagag tgatcggcaa gaaccggcag 1020
cccaagtttg aggaccgggc caagatgccc tacatggagg cagtgatcca cgagatccaa 1080
agatttggag acgtgatccc catgagtttg gcccgcagag tcaaaaagga caccaagttt 1140
cgggatttct tcctccctaa gggcaccgaa gtgtacccta tgctgggctc tgtgctgaga 1200
gaccccagtt tcttctccaa cccccaggac ttcaatcccc agcacttcct gaatgagaag 1260
gggcagttta agaagagtga tgcttttgtg ccctttttcca tcggaaagcg gaactgtttc 1320
ggagaaggcc tggccagaat ggagctcttt ctcttcttca ccaccgtcat gcagaacttc 1380
cgcctcaagt cctcccagtc acctaaggac attgacgtgt cccccaaaca cgtgggcttt 1440
gccacgatcc cacgaaacta caccatgagc ttcctgcccc gctgagtcga cctgcagccc 1500
aagcttccgg tctagaaata attttgttta actttaagaa ggagatataa atatggctga 1560
ctcccacgtg gacaccagct ccaccgtgtc cgaggcggtg gccgaagaag tatctctttt 1620
cagcatgacg gacatgattc tgttttcgct catcgtgggt ctcctaacct actggttcct 1680
cttcagaaag aaaaaagaag aagtccccga gttcaccaaa attcagacat tgacctcctc 1740
tgtcagagag agcagctttg tggaaaagat gaagaaaacg gggaggaaca tcatcgtgtt 1800
ctacggctcc cagacgggga ctgcagagga gtttgccaac cgcctgtcca aggacgccca 1860
ccgctacggg atgcgaggca tgtcagcgga ccctgaggag tatgacctgg ccgacctgag 1920
cagcctgcca gagatcgaca acgccctggt ggttttctgc atggccacct acggtgaggg 1980
agaccccacc gacaatgccc aggacttcta cgactggctg caggagacag acgtggatct 2040
ctctggggtc aagttcgcgg tgtttggtct tgggaacaag acctacgagc acttcaatgc 2100
catgggcaag tacgtggaca agcggctgga gcagctcggc gcccagcgca tctttgagct 2160
ggggttgggc gacgacgatg ggaacttgga ggaggacttc atcacctggc gagagcagtt 2220
ctggccggcc gtgtgtgaac actttggggt ggaagccact ggcgaggagt ccagcattcg 2280
ccagtacgag cttgtggtcc acaccgacat agatgcggcc aaggtgtaca tgggggagat 2340
gggccggctg aagagctacg agaaccagaa gccccccttt gatgccaaga atccgttcct 2400
ggctgcagtc accaccaacc ggaagctgaa ccagggaacc gagcgccacc tcatgcacct 2460
ggaattggac atctcggact ccaaaatcag gtatgaatct ggggaccacg tggctgtgta 2520
cccagccaac gactctgctc tcgtcaacca gctgggcaaa atcctgggtg ccgacctgga 2580
cgtcgtcatg tccctgaaca acctggatga ggagtccaac aagaagcacc cattcccgtg 2640
ccctacgtcc taccgcacgg ccctcaccta ctacctggac atcaccaacc cgccgcgtac 2700
caacgtgctg tacgagctgg cgcagtacgc ctcggagccc tcggagcagg agctgctgcg 2760
caagatggcc tcctcctccg gcgagggcaa ggagctgtac ctgagctggg tggtggaggc 2820

ccggaggcac atcctggcca tcctgcagga ctgcccgtcc ctgcggcccc ccatcgacca 2880
cctgtgtgag ctgctgccgc gcctgcaggc ccgctactac tccatcgcct catcctccaa 2940
ggtccacccc aactctgtgc acatctgtgc ggtggttgtg gagtacgaga ccaaggccgg 3000
ccgcatcaac aagggcgtgg ccaccaactg gctgcgggcc aaggagcctg ccggggagaa 3060
cggcggccgt gcgctggtgc ccatgttcgt gcgcaagtcc cagttccgcc tgcccttcaa 3120
ggccaccacg cctgtcatca tggtgggccc cggcaccggg gtggcaccct tcataggctt 3180
catccaggag cgggcctggc tgcgacagca gggcaaggag gtgggggaga cgctgctgta 3240
ctacggctgc cgccgctcgg atgaggacta cctgtaccgg gaggacgtgg cgcagttcca 3300
cagggacggt gcgctcaccc agctcaacgt ggccttctcc cgggagcagt cccacaaggt 3360
ctacgtccag cacctgctaa agcaagaccg agagcacctg tggaagttga tcgaaggcgg 3420
tgcccacatc tacgtctgtg gggatgcacg gaacatggcc agggatgtgc agaacacctt 3480
ctacgacatc gtggctgagc tcggggccat ggagcacgcg caggcggtgg actacatcaa 3540
gaaactgatg accaagggcc gctactccct ggacgtgtgg agctagaagc ttatcgatga 3600
taagctgtca aacatgagca gatctgagcc cgcctaatga gcgggctttt ttttcagatc 3660
tgcttgaaga cgaaagggcc tcgtgatacg cctattttta taggttaatg tcatgataat 3720
aatggtttct tagcgtcaaa gcaaccatag tacgcgccct gtagcggcgc attaagcgcg 3780
gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct 3840
cctttcgctt tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta 3900
aatcgggggc tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa 3960
cttgatttgg gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct 4020
ttgacgttgg agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc 4080
aaccctatct cgggctattc ttttgattta taagggattt tgccgatttc ggcctattgg 4140
ttaaaaaatg agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgttt 4200
acaatttcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 4260
taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa 4320
tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt 4380
gcggcatttt gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct 4440
gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc 4500
cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta 4560
tgtggcgcgg tattatcccg tgttgacgcc gggcaagagc aactcggtcg ccgcatacac 4620
tattctcaga atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc 4680
atgacagtaa gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac 4740
ttacttctga caacgatcgg aggaccgaag gagctaaccg ctttttgca caacatgggg 4800
gatcatgtaa ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac 4860
gagcgtgaca ccacgatgcc tgcagcaatg gcaacaacgt tgcgcaaact attaactggc 4920
gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt 4980
gcaggaccac ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga 5040
gccggtgagc gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc 5100
cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag 5160

atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca 5220
tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc 5280
ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca 5340
gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc 5400
tgcttgcaaa caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta 5460
ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt 5520
ctagtgtagc cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc 5580
gctctgctaa tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg 5640
ttggactcaa gacgatagtt accggataag gcgcagcggt cgggctgaac gggggttcg 5700
tgcacacagc ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag 5760
ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc 5820
agggtcggaa caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat 5880
agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg 5940
gggcggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc 6000
tggccttttg ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt 6060
accgcctttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca 6120
gtgagcgagg aagcggaaga gcgcctgatg cggtattttc tccttacgca tctgtgcggt 6180
atttcacacc gcatatatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa 6240
gccagtatac actccgctat cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc 6300
aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt acagacaagc 6360
tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc 6420
gaggcagaac gccatcaaaa ataattcgcg tctggccttc ctgtagccag ctttcatcaa 6480
cattaaatgt gagcgagtaa caacccgtcg gattctccgt gggaacaaac ggcggattga 6540
ccgtaatggg ataggttacg ttggtgtaga tgggcgcatc gtaaccgtgc atctgccagt 6600
ttgaggggac gacgacagta tcggcctcag gaagatcgca ctccagccag ctttccggca 6660
ccgcttctgg tgccggaaac caggcaaagc gccattcgcc attcaggctg cgcaactgtt 6720
gggaagggcg atcggtgcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg 6780
ctgcaaggcg attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 6840
cggccagtga atccgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc 6900
acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga 6960
gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg 7020
tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg 7080
cgccagggtg gtttttcttt tcaccagtga gacgggcaac agctgattgc ccttcaccgc 7140
ctggccctga gagagttgca gcaagcggtc cacgctggtt tgccccagca ggcgaaaatc 7200
ctgtttgatg gtggttaacg gcgggatata acatgagctg tcttcggtat cgtcgtatcc 7260
cactaccgag atatccgcac caacgcgcag cccggactcg gtaatggcgc gcattgcgcc 7320
cagcgccatc tgatcgttgg caaccagcat cgcagtggga acgatgccct cattcagcat 7380
ttgcatggtt tgttgaaaac cggacatggc actccagtcg ccttcccgtt ccgctatcgg 7440
ctgaatttga ttgcgagtga gatatttatg ccagccagcc agacgcagac gcgccgagac 7500

agaacttaat gggcccgcta acagcgcgat ttgctggtga cccaatgcga ccagatgctc 7560
cacgcccagt cgcgtaccgt cttcatggga gaaaataata ctgttgatgg gtgtctggtc 7620
agagacatca agaaataacg ccggaacatt agtgcaggca gcttccacag caatggcatc 7680
ctggtcatcc agcggatagt taatgatcag cccactgacg cgttgcgcga gaagattgtg 7740
caccgccgct ttacaggctt cgacgccgct tcgttctacc atcgacacca ccacgctggc 7800
acccagttga tcggcgcgag atttaatcgc cgcgacaatt tgcgacggcg cgtgcagggc 7860
cagactggag gtggcaacgc caatcagcaa cgactgtttg cccgccagtt gttgtgccac 7920
gcggttggga atgtaattca gctccgccat cgccgcttcc acttttccc gcgttttcgc 7980
agaaacgtgg ctggcctggt tcaccacgcg ggaaacggtc tgataagaga caccggcata 8040
ctctgcgaca tcgtataacg ttactggttt cacattcacc accctgaatt gactctcttc 8100
cgggcgctat catgccatac cgcgaaaggt tttgcgccat tcgatggtgt cctggcacga 8160
caggtttccc gactggaaag cgggcagtga gcgcaacgca attaatgtga gttagctcac 8220
tcattaggca ccccaggctt tacactttat gcttccggct cgtataatgt gtggaattgt 8280
gagcggataa caatttcaca caggaaacag gatcgatcca tcgatgagct tactccccat 8340
ccccctgttg acaattaatc atcggctcgt ataatgtgtg gaattgtgag cggataacaa 8400
tttcacacag gaaacaggat cagcttactc cccatccccc tgttgacaat taatcatcgg 8460
ctcgtataat gtgtggaatt gtgagcggat aacaatttca cacaggaaac aggatccatc 8520
gatgcttagg aggtcat 8537

<210> 18
<211> 8492
<212> DNA
<213> Human cells

<400> 18
atggctcgtc aagttcattc ttcttggaat ctgcccccag gccctttccc gcttcccatc 60
atcgggaacc tcttccagtt ggaattgaag aatattccca agtccttcac ccggttggcc 120
cagcgcttcg ggccggtgtt cacgctgtac gtgggctcgc agcgcatggt ggtgatgcac 180
ggctacaagg cggtgaagga agcgctgctg gactacaagg acgagttctc gggcagaggc 240
gacctccccg cgttccatgc gcacagggac aggggaatca tttttaataa tggacctacc 300
tggaaggaca tccggcggtt ttccctgacc accctccgga actatgggat ggggaaacag 360
ggcaatgaga gccggatcca gagggaggcc cacttcctgc tggaagcact caggaagacc 420
caaggccagc ctttcgaccc caccttcctc atcggctgcg cgccctgcaa cgtcatagcc 480
gacatcctct tccgcaagca ttttgactac aatgatgaga agtttctaag gctgatgtat 540
ttgtttaatg agaacttcca cctactcagc actccctggc tccagcttta caataatttt 600
cccagctttc tacactactt gcctggaagc cacagaaaag tcataaaaaa tgtggctgaa 660
gtaaaagagt atgtgtctga aagggtgaag gagcaccatc aatctctgga ccccaactgt 720
ccccgggacc tcaccgactg cctgctcgtg gaaatggaga aggaaaagca cagtgcagag 780

cgcttgtaca caatggacgg tatcaccgtg actgtggccg acctgttctt tgcggggaca 840
gagaccacca gcacaactct gagatatggg ctcctgattc tcatgaaata ccctgagatc 900
gaagagaagc tccatgaaga aattgacagg gtgattgggc caagccgaat ccctgccatc 960
aaggataggc aagagatgcc ctacatggat gctgtggtgc atgagattca gcggttcatc 1020
accctcgtgc cctccaacct gccccatgaa gcaacccgag acaccatttt cagaggatac 1080
ctcatcccca agggcacagt cgtagtgcca actctggact ctgttttgta tgacaaccaa 1140
gaatttcctg atccagaaaa gtttaagcca gaacacttcc tgaatgaaaa tggaaagttc 1200
aagtacagtg actatttcaa gccattttcc acaggaaaac gagtgtgtgc tggagaaggc 1260
ctggctcgca tggagttgtt tcttttgttg tgtgccattt tgcagcattt taatttgaag 1320
cctctcgttg acccaaagga tatcgacctc agccctatac atattgggtt tggctgtatc 1380
ccaccacgtt acaaactctg tgtcattccc cgctcatgag tgtgtggagg acaccctgaa 1440
ccccccgctt caaacaagct tccggtctag aaataatttt gtttaacttt aagaaggaga 1500
tataaatatg gctgactccc acgtggacac cagctccacc gtgtccgagg cggtggccga 1560
agaagtatct cttttcagca tgacggacat gattctgttt tcgctcatcg tgggtctcct 1620
aacctactgg ttcctcttca gaaagaaaaa agaagaagtc cccgagttca ccaaaattca 1680
gacattgacc tcctctgtca gagagagcag ctttgtggaa aagatgaaga aaacggggag 1740
gaacatcatc gtgttctacg gctcccagac ggggactgca gaggagtttg ccaaccgcct 1800
gtccaaggac gcccaccgct acgggatgcg aggcatgtca gcggaccctg aggagtatga 1860
cctggccgac ctgagcagcc tgccagagat cgacaacgcc ctggtggttt tctgcatggc 1920
cacctacggt gagggagacc ccaccgacaa tgcccaggac ttctacgact ggctgcagga 1980
gacagacgtg gatctctctg gggtcaagtt cgcggtgttt ggtcttggga acaagaccta 2040
cgagcacttc aatgccatgg gcaagtacgt ggacaagcgg ctggagcagc tcggcgccca 2100
gcgcatcttt gagctggggt tgggcgacga cgatgggaac ttggaggagg acttcatcac 2160
ctggcgagag cagttctggc cggccgtgtg tgaacacttt ggggtggaag ccactggcga 2220
ggagtccagc attcgccagt acgagcttgt ggtccacacc gacatagatg cggccaaggt 2280
gtacatgggg gagatgggcc ggctgaagag ctacgagaac cagaagcccc cctttgatgc 2340
caagaatccg ttcctggctg cagtcaccac caaccggaag ctgaaccagg gaaccgagcg 2400
ccacctcatg cacctggaat tggacatctc ggactccaaa atcaggtatg aatctgggga 2460
ccacgtggct gtgtacccag ccaacgactc tgctctcgtc aaccagctgg gcaaaatcct 2520
gggtgccgac ctggacgtcg tcatgtccct gaacaacctg gatgaggagt ccaacaagaa 2580
gcacccattc ccgtgcccta cgtcctaccg cacggccctc acctactacc tggacatcac 2640
caacccgccg cgtaccaacg tgctgtacga gctggcgcag tacgcctcgg agccctcgga 2700
gcaggagctg ctgcgcaaga tggcctcctc ctccggcgag ggcaaggagc tgtacctgag 2760
ctgggtggtg gaggcccgga ggcacatcct ggccatcctg caggactgcc cgtccctgcg 2820
gccccccatc gaccacctgt gtgagctgct gccgcgcctg caggcccgct actactccat 2880
cgcctcatcc tccaaggtcc accccaactc tgtgcacatc tgtgcggtgg ttgtggagta 2940
cgagaccaag gccggccgca tcaacaaggg cgtggccacc aactggctgc gggccaagga 3000
gcctgccggg gagaacggcg gccgtgcgct ggtgcccatg ttcgtgcgca agtcccagtt 3060
ccgcctgccc ttcaaggcca ccacgcctgt catcatggtg ggccccggca ccggggtggc 3120

acccttcata ggcttcatcc aggagcgggc ctggctgcga cagcagggca aggaggtggg 3180
ggagacgctg ctgtactacg gctgccgccg ctcggatgag gactacctgt accgggagga 3240
cgtggcgcag ttccacaggg acggtgcgct cacccagctc aacgtggcct tctcccggga 3300
gcagtcccac aaggtctacg tccagcacct gctaaagcaa gaccgagagc acctgtggaa 3360
gttgatcgaa ggcggtgccc acatctacgt ctgtggggat gcacggaaca tggccaggga 3420
tgtgcagaac accttctacg acatcgtggc tgagctcggg gccatggagc acgcgcaggc 3480
ggtggactac atcaagaaac tgatgaccaa gggccgctac tccctggacg tgtggagcta 3540
gaagcttatc gatgataagc tgtcaaacat gagcagatct gagcccgcct aatgagcggg 3600
ctttttttc agatctgctt gaagacgaaa gggcctcgtg atacgcctat ttttataggt 3660
taatgtcatg ataataatgg tttcttagcg tcaaagcaac catagtacgc gccctgtagc 3720
ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg tgaccgctac acttgccagc 3780
gccctagcgc ccgctccttt cgctttcttc ccttcctttc tcgccacgtt cgccggcttt 3840
ccccgtcaag ctctaaatcg ggggctccct ttagggttcc gatttagtgc tttacggcac 3900
ctcgacccca aaaaacttga tttgggtgat ggttcacgta gtgggccatc gccctgatag 3960
acggtttttc gccctttgac gttggagtcc acgttcttta atagtggact cttgttccaa 4020
actggaacaa cactcaaccc tatctcgggc tattcttttg atttataagg gattttgccg 4080
atttcggcct attggttaaa aaatgagctg atttaacaaa aatttaacgc gaatttttaac 4140
aaaatattaa cgtttacaat ttcaggtggc acttttcggg gaaatgtgcg cggaacccct 4200
atttgtttat ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga 4260
taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc 4320
cttattccct tttttgcggc attttgcctt cctgtttttg ctcacccaga aacgctggtg 4380
aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg gttacatcga actggatctc 4440
aacagcggta agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact 4500
tttaaagttc tgctatgtgg cgcggtatta tcccgtgttg acgccgggca agagcaactc 4560
ggtcgccgca tacactattc tcagaatgac ttggttgagt actcaccagt cacagaaaag 4620
catcttacgg atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat 4680
aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt 4740
ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa 4800
gccataccaa acgacgagcg tgacaccacg atgcctgcag caatggcaac aacgttgcgc 4860
aaactattaa ctggcgaact acttactcta gcttcccggc aacaattaat agactggatg 4920
gaggcggata aagttgcagg accacttctg cgctcggccc ttccggctgg ctggtttatt 4980
gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcattgcagc actggggcca 5040
gatggtaagc cctcccgtat cgtagttatc tacacgacgg ggagtcaggc aactatggat 5100
gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca 5160
gaccaagttt actcatatat actttagatt gatttaaaac ttcattttta atttaaaagg 5220
atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg 5280
ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt 5340
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg 5400
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 5460

ccaaatactg tccttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 5520
ccgcctacat acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag 5580
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 5640
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 5700
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg 5760
tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac 5820
gcctggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 5880
tgatgctcgt caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 5940
ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcgttatc ccctgattct 6000
gtggataacc gtattaccgc ctttgagtga gctgataccg ctcgccgcag ccgaacgacc 6060
gagcgcagcg agtcagtgag cgaggaagcg gaagagcgcc tgatgcggta ttttctcctt 6120
acgcatctgt gcggtatttc acaccgcata tatggtgcac tctcagtaca atctgctctg 6180
atgccgcata gttaagccag tatacactcc gctatcgcta cgtgactggg tcatggctgc 6240
gccccgacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc tcccggcatc 6300
cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtcagaggt tttcaccgtc 6360
atcaccgaaa cgcgcgaggc agaacgccat caaaaataat tcgcgtctgg ccttcctgta 6420
gccagctttc atcaacatta aatgtgagcg agtaacaacc cgtcggattc tccgtgggaa 6480
caaacggcgg attgaccgta atgggatagg ttacgttggt gtagatgggc gcatcgtaac 6540
cgtgcatctg ccagtttgag gggacgacga cagtatcggc ctcaggaaga tcgcactcca 6600
gccagctttc cggcaccgct tctggtgccg gaaaccaggc aaagcgccat tcgccattca 6660
ggctgcgcaa ctgttgggaa gggcgatcgg tgcgggcctc ttcgctatta cgccagctgg 6720
cgaaaggggg atgtgctgca aggcgattaa gttgggtaac gccagggttt tcccagtcac 6780
gacgttgtaa aacgacggcc agtgaatccg taatcatggt catagctgtt tcctgtgtga 6840
aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc 6900
tggggtgcct aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc 6960
cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc 7020
ggtttgcgta ttgggcgcca gggtggtttt tcttttcacc agtgagacgg gcaacagctg 7080
attgcccttc accgcctggc cctgagagag ttgcagcaag cggtccacgc tggtttgccc 7140
cagcaggcga aaatcctgtt tgatggtggt taacggcggg atataacatg agctgtcttc 7200
ggtatcgtcg tatcccacta ccgagatatc cgcaccaacg cgcagcccgg actcggtaat 7260
ggcgcgcatt gcgcccagcg ccatctgatc gttggcaacc agcatcgcag tgggaacgat 7320
gccctcattc agcatttgca tggtttgttg aaaaccggac atggcactcc agtcgccttc 7380
ccgttccgct atcggctgaa tttgattgcg agtgagatat ttatgccagc cagccagacg 7440
cagacgcgcc gagacagaac ttaatgggcc cgctaacagc gcgatttgct ggtgacccaa 7500
tgcgaccaga tgctccacgc ccagtcgcgt accgtcttca tgggagaaaa taatactgtt 7560
gatgggtgtc tggtcagaga catcaagaaa taacgccgga acattagtgc aggcagcttc 7620
cacagcaatg gcatcctggt catccagcgg atagttaatg atcagcccac tgacgcgttg 7680
cgcgagaaga ttgtgcaccg ccgctttaca ggcttcgacg ccgcttcgtt ctaccatcga 7740
caccaccacg ctggcaccca gttgatcggc gcgagattta atcgccgcga caatttgcga 7800

cggcgcgtgc agggccagac tggaggtggc aacgccaatc agcaacgact gtttgcccgc 7860
cagttgttgt gccacgcggt tgggaatgta attcagctcc gccatcgccg cttccacttt 7920
ttcccgcgtt ttcgcagaaa cgtggctggc ctggttcacc acgcgggaaa cggtctgata 7980
agagacaccg gcatactctg cgacatcgta taacgttact ggtttcacat tcaccaccct 8040
gaattgactc tcttccgggc gctatcatgc cataccgcga aaggttttgc gccattcgat 8100
ggtgtcctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa 8160
tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat 8220
aatgtgtgga attgtgagcg gataacaatt tcacacagga aacaggatcg atccatcgat 8280
gagcttactc cccatccccc tgttgacaat taatcatcgg ctcgtataat gtgtggaatt 8340
gtgagcggat aacaatttca cacaggaaac aggatcagct tactccccat ccccctgttg 8400
acaattaatc atcggctcgt ataatgtgtg gaattgtgag cggataacaa tttcacacag 8460
gaaacaggat ccatcgatgc ttaggaggtc at 8492

<210> 19
<211> 8475
<212> DNA
<213> Human cells

<400> 19
atggctcgac aatcttctgg acgaggaaaa ctccctcctg gccctactcc tctcccagtg 60
attggaaata tcctacagat agatattaag gatgtcagca aatccttaac caatctctca 120
aaaatctatg gccctgtgtt cactctgtat tttggcctgg aacgcatggt ggtgctgcat 180
ggatatgaag tggtgaagga agccctgatt gatcttggag aggagtttc tggaagaggc 240
catttcccac tggctgaaag agctaacaga ggatttggaa tcgttttcag caatggaaag 300
agatggaagg agatccggcg tttctccctc atgacgctgc ggaattttgg gatggggaag 360
aggagcattg aggaccgtgt tcaagaggaa gcccgctgcc ttgtggagga gttgagaaaa 420
accaaggctt caccctgtga tcccactttc atcctgggct gtgctccctg caatgtgatc 480
tgctccatta ttttccagaa acgtttcgat tataaagatc agcaatttct taacttgatg 540
gaaaaattga atgaaaacat caggattgta agcacccccct ggatccagat atgcaataat 600
tttcccacta tcattgatta tttcccggga acccataaca aattacttaa aaaccttgct 660
tttatggaaa gtgatatttt ggagaaagta aaagaacacc aagaatcgat ggacatcaac 720
aaccctcggg actttattga ttgcttcctg atcaaaatgg agaaggaaaa gcaaaaccaa 780
cagtctgaat tcactattga aaacttggta atcactgcag ctgacttact tggagctggg 840
acagagacaa caagcacaac cctgagatat gctctccttc tcctgctgaa gcacccagag 900
gtcacagcta aagtccagga agagattgaa cgtgtcgttg gcagaaaccg gagcccctgc 960
atgcaggaca ggggccacat gccctacaca gatgctgtgg tgcacgaggt ccagagatac 1020
atcgacctca tccccaccag cctgccccat gcagtgacct gtgacgttaa attcagaaac 1080
tacctcattc ccaagggcac aaccatatta acttccctca cttctgtgct acatgacaac 1140

aaagaatttc ccaacccaga gatgtttgac cctcgtcact ttctggatga aggtggaaat 1200
tttaagaaaa gtaactactt catgcctttc tcagcaggaa aacggatttg tgtgggagag 1260
ggcctggccc gcatggagct gtttttattc ctgaccttca ttttacagaa ctttaacctg 1320
aaatctctga ttgacccaaa ggaccttgac acaactcctg ttgtcaatgg atttgcttct 1380
gtcccgccct tctatcagct gtgcttcatt cctgtctgaa gaagcacaga tggtctggaa 1440
gcttccggtc tagaaataat tttgtttaac tttaagaagg agatataaat atggctgact 1500
cccacgtgga caccagctcc accgtgtccg aggcggtggc cgaagaagta tctcttttca 1560
gcatgacgga catgattctg ttttcgctca tcgtgggtct cctaacctac tggttcctct 1620
tcagaaagaa aaaagaagaa gtccccgagt tcaccaaaat tcagacattg acctcctctg 1680
tcagagagag cagctttgtg gaaaagatga agaaaacggg gaggaacatc atcgtgttct 1740
acggctccca gacggggact gcagaggagt ttgccaaccg cctgtccaag gacgcccacc 1800
gctacgggat gcgaggcatg tcagcggacc ctgaggagta tgacctggcc gacctgagca 1860
gcctgccaga gatcgacaac gccctggtgg ttttctgcat ggccacctac ggtgagggag 1920
accccaccga caatgcccag gacttctacg actggctgca ggagacagac gtggatctct 1980
ctggggtcaa gttcgcggtg tttggtcttg ggaacaagac ctacgagcac ttcaatgcca 2040
tgggcaagta cgtggacaag cggctggagc agctcggcgc ccagcgcatc tttgagctgg 2100
ggttgggcga cgacgatggg aacttggagg aggacttcat cacctggcga gagcagttct 2160
ggccggccgt gtgtgaacac tttggggtgg aagccactgg cgaggagtcc agcattcgcc 2220
agtacgagct tgtggtccac accgacatag atgcggccaa ggtgtacatg ggggagatgg 2280
gccggctgaa gagctacgag aaccagaagc ccccctttga tgccaagaat ccgttcctgg 2340
ctgcagtcac caccaaccgg aagctgaacc agggaaccga gcgccacctc atgcacctgg 2400
aattggacat ctcggactcc aaaatcaggt atgaatctgg ggaccacgtg gctgtgtacc 2460
cagccaacga ctctgctctc gtcaaccagc tgggcaaaat cctgggtgcc gacctggacg 2520
tcgtcatgtc cctgaacaac ctggatgagg agtccaacaa gaagcaccca ttcccgtgcc 2580
ctacgtccta ccgcacggcc ctcacctact acctggacat caccaacccg ccgcgtacca 2640
acgtgctgta cgagctggcg cagtacgcct cggagccctc ggagcaggag ctgctgcgca 2700
agatggcctc ctcctccggc gagggcaagg agctgtacct gagctgggtg gtggaggccc 2760
ggaggcacat cctggccatc ctgcaggact gcccgtccct gcggcccccc atcgaccacc 2820
tgtgtgagct gctgccgcgc ctgcaggccc gctactactc catcgcctca tcctccaagg 2880
tccaccccaa ctctgtgcac atctgtgcgg tggttgtgga gtacgagacc aaggccggcc 2940
gcatcaacaa gggcgtggcc accaactggc tgcgggccaa ggagcctgcc ggggagaacg 3000
gcggccgtgc gctggtgccc atgttcgtgc gcaagtccca gttccgcctg cccttcaagg 3060
ccaccacgcc tgtcatcatg gtgggccccg gcaccggggt ggcaccctte ataggcttca 3120
tccaggagcg ggcctggctg cgacagcagg gcaaggaggt gggggagacg ctgctgtact 3180
acggctgccg ccgctcggat gaggactacc tgtaccggga ggacgtggcg cagttccaca 3240
gggacggtgc gctcacccag ctcaacgtgg ccttctcccg ggagcagtcc cacaaggtct 3300
acgtccagca cctgctaaag caagaccgag agcacctgtg gaagttgatc gaaggcggtg 3360
cccacatcta cgtctgtggg gatgcacgga acatggccag ggatgtgcag aacaccttct 3420
acgacatcgt ggctgagctc ggggccatgg agcacgcgca ggcggtggac tacatcaaga 3480

aactgatgac caagggccgc tactccctgg acgtgtggag ctagaagctt atcgatgata 3540
agctgtcaaa catgagcaga tctgagcccg cctaatgagc gggctttttt ttcagatctg 3600
cttgaagacg aaagggcctc gtgatacgcc tatttttata ggttaatgtc atgataataa 3660
tggtttctta gcgtcaaagc aaccatagta cgcgccctgt agcggcgcat taagcgcggc 3720
gggtgtggtg gttacgcgca gcgtgaccgc tacacttgcc agcgccctag cgcccgctcc 3780
tttcgctttc ttcccttcct ttctcgccac gttcgccggc tttccccgtc aagctctaaa 3840
tcgggggctc cctttagggt tccgatttag tgctttacgg cacctcgacc ccaaaaaact 3900
tgatttgggt gatggttcac gtagtgggcc atcgccctga tagacggttt ttcgcccttt 3960
gacgttggag tccacgttct ttaatagtgg actcttgttc caaactggaa caacactcaa 4020
ccctatctcg ggctattctt ttgatttata agggattttg ccgatttcgg cctattggtt 4080
aaaaaatgag ctgatttaac aaaaatttaa cgcgaatttt aacaaaatat taacgtttac 4140
aatttcaggt ggcacttttc ggggaaatgt gcgcggaacc cctatttgtt tatttttcta 4200
aatacattca aatatgtatc cgctcatgag acaataaccc tgataaatgc ttcaataata 4260
ttgaaaaagg aagagtatga gtattcaaca tttccgtgtc gcccttattc cctttttgc 4320
ggcattttgc cttcctgttt ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga 4380
agatcagttg ggtgcacgag tgggttacat cgaactggat ctcaacagcg gtaagatcct 4440
tgagagtttt cgccccgaag aacgttttcc aatgatgagc acttttaaag ttctgctatg 4500
tggcgcggta ttatcccgtg ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta 4560
ttctcagaat gacttggttg agtactcacc agtcacagaa aagcatctta cggatggcat 4620
gacagtaaga gaattatgca gtgctgccat aaccatgagt gataacactg cggccaactt 4680
acttctgaca acgatcggag gaccgaagga gctaaccgct tttttgcaca acatgggggga 4740
tcatgtaact cgccttgatc gttgggaacc ggagctgaat gaagccatac caaacgacga 4800
gcgtgacacc acgatgcctg cagcaatggc aacaacgttg cgcaaactat taactggcga 4860
actacttact ctagcttccc ggcaacaatt aatagactgg atggaggcgg ataaagttgc 4920
aggaccactt ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc 4980
cggtgagcgt gggtctcgcg gtatcattgc agcactgggg ccagatggta agccctcccg 5040
tatcgtagtt atctacacga cggggagtca ggcaactatg gatgaacgaa atagacagat 5100
cgctgagata ggtgcctcac tgattaagca ttggtaactg tcagaccaag tttactcata 5160
tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg tgaagatcct 5220
ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact gagcgtcaga 5280
ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg taatctgctg 5340
cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc aagagctacc 5400
aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata ctgtccttct 5460
agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta catacctcgc 5520
tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt 5580
ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg 5640
cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac agcgtgagct 5700
atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg taagcggcag 5760
ggtcggaaca ggagagcgca cgagggagct tccaggggga aacgcctggt atctttatag 5820

tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg 5880
gcggagccta tggaaaaacg ccagcaacgc ggccttttta cggttcctgg ccttttgctg 5940
gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata accgtattac 6000
cgcctttgag tgagctgata ccgctcgccg cagccgaacg accgagcgca gcgagtcagt 6060
gagcgaggaa gcggaagagc gcctgatgcg gtattttctc cttacgcatc tgtgcggtat 6120
ttcacaccgc atatatggtg cactctcagt acaatctgct ctgatgccgc atagttaagc 6180
cagtatacac tccgctatcg ctacgtgact gggtcatggc tgcgccccga cacccgccaa 6240
cacccgctga cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg 6300
tgaccgtctc cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga 6360
ggcagaacgc catcaaaaat aattcgcgtc tggccttcct gtagccagct ttcatcaaca 6420
ttaaatgtga gcgagtaaca acccgtcgga ttctccgtgg gaacaaacgg cggattgacc 6480
gtaatgggat aggttacgtt ggtgtagatg ggcgcatcgt aaccgtgcat ctgccagttt 6540
gaggggacga cgacagtatc ggcctcagga agatcgcact ccagccagct ttccggcacc 6600
gcttctggtg ccggaaacca ggcaaagcgc cattcgccat tcaggctgcg caactgttgg 6660
gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct 6720
gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg taaaacgacg 6780
gccagtgaat ccgtaatcat ggtcatagct gtttcctgtg tgaaattgtt atccgctcac 6840
aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt 6900
gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 6960
gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 7020
ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 7080
ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 7140
gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 7200
ctaccgagat atccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 7260
gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 7320
gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 7380
gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 7440
aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 7500
cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 7560
agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 7620
ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 7680
ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 7740
ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 7800
gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 7860
ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 7920
aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 7980
ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 8040
ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc tggcacgaca 8100
ggtttcccga ctggaaagcg ggcagtgagc gcaacgcaat taatgtgagt tagctcactc 8160

attaggcacc ccaggcttta cactttatgc ttccggctcg tataatgtgt ggaattgtga 8220
gcggataaca atttcacaca ggaaacagga tcgatccatc gatgagctta ctccccatcc 8280
ccctgttgac aattaatcat cggctcgtat aatgtgtgga attgtgagcg gataacaatt 8340
tcacacagga aacaggatca gcttactccc catccccctg ttgacaatta atcatcggct 8400
cgtataatgt gtggaattgt gagcggataa caatttcaca caggaaacag gatccatcga 8460
tgcttaggag gtcat 8475

<210> 20
<211> 1215
<212> DNA
<213> Human cells

<220>
<221> CDS
<222> (1)..(1212)

<400> 20
atg act gcg aca gtc gct tcg aca agt ctg ttc acc acc gcc gac cac 48
Met Thr Ala Thr Val Ala Ser Thr Ser Leu Phe Thr Thr Ala Asp His
1 5 10 15

tac cac act ccg ttg ggt ccc gac ggc acg cca cac gca ttc ttc gag 96
Tyr His Thr Pro Leu Gly Pro Asp Gly Thr Pro His Ala Phe Phe Glu
20 25 30

gcg ttg cgc gac gag gcg gag acc acc cct att gga tgg agt gag gcc 144
Ala Leu Arg Asp Glu Ala Glu Thr Thr Pro Ile Gly Trp Ser Glu Ala
35 40 45

tac ggt ggc cac tgg gta gtt gcc gga tac aag gag att cag gct gtc 192
Tyr Gly Gly His Trp Val Val Ala Gly Tyr Lys Glu Ile Gln Ala Val
50 55 60

atc cag aac acc aag gcc ttc tcg aac aag ggc gtg acc ttc cct cgc 240
Ile Gln Asn Thr Lys Ala Phe Ser Asn Lys Gly Val Thr Phe Pro Arg
65 70 75 80

tac gaa acc ggc gag ttc gag ctc atg atg gca ggg cag gac gac ccc 288

```
Tyr Glu Thr Gly Glu Phe Glu Leu Met Met Ala Gly Gln Asp Asp Pro
                 85                  90                  95

gtc cac aag aag tac cgc cag ctc gtg gcg aaa ccg ttc tcg ccg gag   336
Val His Lys Lys Tyr Arg Gln Leu Val Ala Lys Pro Phe Ser Pro Glu
            100                 105                 110

gcg acc gac ctg ttc acc gaa cag ttg cgc cag tcc acc aac gac ctg   384
Ala Thr Asp Leu Phe Thr Glu Gln Leu Arg Gln Ser Thr Asn Asp Leu
        115                 120                 125

atc gac gcg cgg atc gaa ctc gga gag ggt gac gct gcc acc tgg tta   432
Ile Asp Ala Arg Ile Glu Leu Gly Glu Gly Asp Ala Ala Thr Trp Leu
    130                 135                 140

gcc aac gag att cct gct cgg cta acg gcc atc ctc ctc ggc ctt ccg   480
Ala Asn Glu Ile Pro Ala Arg Leu Thr Ala Ile Leu Leu Gly Leu Pro
145                 150                 155                 160

ccc gaa gac gga gac acc tac cgc cgg tgg gtg tgg gcc atc acg cac   528
Pro Glu Asp Gly Asp Thr Tyr Arg Arg Trp Val Trp Ala Ile Thr His
                165                 170                 175

gta gag aac ccg gag gaa ggc gcc gaa att ttc gcc gaa ctt gtg gcg   576
Val Glu Asn Pro Glu Glu Gly Ala Glu Ile Phe Ala Glu Leu Val Ala
            180                 185                 190

cac gcg cgc act ctg atc gcc gag cgc cgc act aac cca ggc aac gac   624
His Ala Arg Thr Leu Ile Ala Glu Arg Arg Thr Asn Pro Gly Asn Asp
        195                 200                 205

atc atg tct cgg gtc atc atg tcc aag atc gac ggt gaa agc ctc tcc   672
Ile Met Ser Arg Val Ile Met Ser Lys Ile Asp Gly Glu Ser Leu Ser
    210                 215                 220

gag gac gac ctg atc gga ttc ttc acc atc ctg ctt ctc ggc ggc atc   720
Glu Asp Asp Leu Ile Gly Phe Phe Thr Ile Leu Leu Leu Gly Gly Ile
225                 230                 235                 240
```

```
gac aac acc gca cgc ttc ctc agc tcg gtg ttc tgg cgc ttg gcg tgg   768
Asp Asn Thr Ala Arg Phe Leu Ser Ser Val Phe Trp Arg Leu Ala Trp
                245                 250                 255

gac atc gag ctc cgc cgc cga ctc atc gcc cac ccc gag ctt atc ccc   816
Asp Ile Glu Leu Arg Arg Arg Leu Ile Ala His Pro Glu Leu Ile Pro
            260                 265                 270

aac gcc gta gac gaa ctg ctc cgc ttc tac ggc ccc gcc atg gtc ggc   864
Asn Ala Val Asp Glu Leu Leu Arg Phe Tyr Gly Pro Ala Met Val Gly
        275                 280                 285

cgc ctc gtc acc caa gag gtc acc gtg ggc gac atc aca atg aag cca   912
Arg Leu Val Thr Gln Glu Val Thr Val Gly Asp Ile Thr Met Lys Pro
    290                 295                 300

ggc caa acg gcc atg ctg tgg ttc ccg atc gcc agc cgg gat cgt tcg   960
Gly Gln Thr Ala Met Leu Trp Phe Pro Ile Ala Ser Arg Asp Arg Ser
305                 310                 315                 320

gca ttc gat tca ccc gac aac atc gtg atc gaa cgc acg ccc aac cgg   1008
Ala Phe Asp Ser Pro Asp Asn Ile Val Ile Glu Arg Thr Pro Asn Arg
                325                 330                 335

cac ctg tct ctc ggc cac ggc atc cac cgc tgc ctc gga gcc cac ctg   1056
His Leu Ser Leu Gly His Gly Ile His Arg Cys Leu Gly Ala His Leu
            340                 345                 350

atc cgc gtc gag gcc cgc gtg gcg atc acc gag ttc ctc aag cgc att   1104
Ile Arg Val Glu Ala Arg Val Ala Ile Thr Glu Phe Leu Lys Arg Ile
        355                 360                 365

ccg gag ttc tcc ctc gac ccg aac aag gaa tgt gag tgg tta atg ggc   1152
Pro Glu Phe Ser Leu Asp Pro Asn Lys Glu Cys Glu Trp Leu Met Gly
    370                 375                 380

cag gtg gcc ggc atg ctg cac gtg ccg atc atc ttc cct aag ggc aag   1200
Gln Val Ala Gly Met Leu His Val Pro Ile Ile Phe Pro Lys Gly Lys
385                 390                 395                 400
```

cga ctg tcg gaa tga 1215
Arg Leu Ser Glu

<210> 21
<211> 404
<212> PRT
<213> Human cells

<400> 21
Met Thr Ala Thr Val Ala Ser Thr Ser Leu Phe Thr Thr Ala Asp His
1 5 10 15

Tyr His Thr Pro Leu Gly Pro Asp Gly Thr Pro His Ala Phe Phe Glu
20 25 30

Ala Leu Arg Asp Glu Ala Glu Thr Thr Pro Ile Gly Trp Ser Glu Ala
35 40 45

Tyr Gly Gly His Trp Val Val Ala Gly Tyr Lys Glu Ile Gln Ala Val
50 55 60

Ile Gln Asn Thr Lys Ala Phe Ser Asn Lys Gly Val Thr Phe Pro Arg
65 70 75 80

Tyr Glu Thr Gly Glu Phe Glu Leu Met Met Ala Gly Gln Asp Asp Pro
85 90 95

Val His Lys Lys Tyr Arg Gln Leu Val Ala Lys Pro Phe Ser Pro Glu
100 105 110

Ala Thr Asp Leu Phe Thr Glu Gln Leu Arg Gln Ser Thr Asn Asp Leu
115 120 125

Ile Asp Ala Arg Ile Glu Leu Gly Glu Gly Asp Ala Ala Thr Trp Leu
130 135 140

Ala Asn Glu Ile Pro Ala Arg Leu Thr Ala Ile Leu Leu Gly Leu Pro

```
145                 150                 155                 160

Pro Glu Asp Gly Asp Thr Tyr Arg Arg Trp Val Trp Ala Ile Thr His
                165                 170                 175

Val Glu Asn Pro Glu Glu Gly Ala Glu Ile Phe Ala Glu Leu Val Ala
            180                 185                 190

His Ala Arg Thr Leu Ile Ala Glu Arg Arg Thr Asn Pro Gly Asn Asp
        195                 200                 205

Ile Met Ser Arg Val Ile Met Ser Lys Ile Asp Gly Glu Ser Leu Ser
    210                 215                 220

Glu Asp Asp Leu Ile Gly Phe Phe Thr Ile Leu Leu Leu Gly Gly Ile
225                 230                 235                 240

Asp Asn Thr Ala Arg Phe Leu Ser Ser Val Phe Trp Arg Leu Ala Trp
                245                 250                 255

Asp Ile Glu Leu Arg Arg Arg Leu Ile Ala His Pro Glu Leu Ile Pro
            260                 265                 270

Asn Ala Val Asp Glu Leu Leu Arg Phe Tyr Gly Pro Ala Met Val Gly
        275                 280                 285

Arg Leu Val Thr Gln Glu Val Thr Val Gly Asp Ile Thr Met Lys Pro
    290                 295                 300

Gly Gln Thr Ala Met Leu Trp Phe Pro Ile Ala Ser Arg Asp Arg Ser
305                 310                 315                 320

Ala Phe Asp Ser Pro Asp Asn Ile Val Ile Glu Arg Thr Pro Asn Arg
                325                 330                 335

His Leu Ser Leu Gly His Gly Ile His Arg Cys Leu Gly Ala His Leu
            340                 345                 350

Ile Arg Val Glu Ala Arg Val Ala Ile Thr Glu Phe Leu Lys Arg Ile
```

355 360 365

Pro Glu Phe Ser Leu Asp Pro Asn Lys Glu Cys Glu Trp Leu Met Gly
370 375 380

Gln Val Ala Gly Met Leu His Val Pro Ile Ile Phe Pro Lys Gly Lys
385 390 395 400

Arg Leu Ser Glu

<210> 23
<211> 154
<212> PRT
<213> Human cells

<400> 23
Met Asn Ala Leu Ile Leu Tyr Gly Thr Glu Thr Gly Asn Ala Glu Ala
1 5 10 15

Cys Ala Thr Thr Ile Ser Gln Val Leu Ala Asp Thr Val Asp Thr Lys
20 25 30

Val His Asp Leu Ala Asp Met Thr Pro Arg Ala Met Leu Asp Ser Gly
35 40 45

Ala Asp Leu Ile Val Phe Ala Thr Ala Thr Tyr Gly Glu Gly Glu Phe
50 55 60

Ala Gly Gly Gly Ala Ala Phe Phe Glu Thr Leu Arg Glu Thr Lys Pro
65 70 75 80

Asp Leu Ser Gly Leu Arg Phe Ala Val Phe Gly Leu Gly Asp Ser Tyr
85 90 95

Tyr Thr Thr Phe Asn Gln Ala Gly Ala Thr Ala Ala Thr Ile Leu Ala
100 105 110

```
Ser Leu Gly Gly Thr Gln Val Gly Asp Thr Ala Arg His Asp Thr Ser
        115                 120                 125

Ser Gly Asp Asp Pro Glu Glu Thr Ala Glu Glu Trp Ala Arg Glu Ile
    130                 135                 140

Leu Thr Ala Leu Ala Thr Pro Ala Val Ser
145                 150
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU00/00966

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: C12N 15/53, 15/63

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC (WPID) AND CHEMICAL ABSTRACTS. KEYWORDS BELOW

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
MEDLINE. KEYWORDS BELOW

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPID, Medline, Chemical Abstracts. Keywords: cytochrome p450 p-450 p4502a6 p4502e1 p4502c19 cyp2a6 cyp2e1 cyp2c19 colour color dye pigment stain indole

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AU 29560/92 (KYOWA HAKKO KOGYO CO., LTD.) 5 October 1993	15-24
X	Eur. J. Biochem., 1996, vol. 239, Munro et al., "Probing electron transfer in flavocytochrome P-450 BM3 and its component domains", pages 403-9	52-61
P,X	FEBS Letters, 1999, vol. 461, Shimada et al., "Expression of chimeric P450 genes encoding flavonoid-3',5'-hydroxylase in transgenic tobacco and petunia plants", pages 241-245	15-24

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 2000	10 OCT 2000
Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	CHRISTOPHER LUTON Telephone No : (02) 6283 2256

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU00/00966

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Drug Metabolism and Drug Interactions, 1995, vol. 12, no. 3-4, Holton, "Modification of flower colour via manipulation of P450 gene expression in transgenic plants", pages 359-368	15-24
X	Nature, 1993, vol. 366, Holton *et al.*, "Cloning and expression of cytochrome P450 genes controlling flower colour", pages 276-9	15-24
X	EP 0522880 (INTERNATIONAL FLOWER DEVELOPMENTS PTY. LTD.) 13 January 1993	15-24
X	Proc. Natl. Acad. Sci. USA, January 1999, vol. 96, Vetten *et al.*, "A cytochrome b5 is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors", pages 778-83	15-24

INTERNATIONÁL SEARCH REPORT

International application No.
PCT/AU00/00966

Box I **Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos :

because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos : **1-14, 25-48**

because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

These claims do not define the matter for which protection is sought in terms limited by the technical features of the invention.

3. ☐ Claims Nos :

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II **Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU00/00966

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member	
AU	29560/92	EP	632128
		US	6114601
		WO	93/18155
EP	522880	AU	19530/92
		AU	22733/92
		AU	67895/94
		CA	2112373
		CA	2163220
		EP	703982
		JP	6500239
		JP	8511683
		JP	2000023686
		US	5349125
		US	5569832
		US	5861487
		US	5948955
		WO	93/01290
		WO	94/28140

END OF ANNEX